V.E. GMURMAN

TEORIA DE LAS PROBABILIDADES Y ESTADISTICA MATEMATICA


ИЗДАТЕЛЬСТВО
МИР

В. Е. Гмурман

Теория вероятностей и математическая статистика

ИЗДАТЕЛЬСТВО «ВЫСШАЯ ШКОЛА»

V. E. Gmurman

Teoría de las probabilidades y estadística matemática

Traducido del ruso por el ingeniero Akop Grdian

MOSCÚ EDITORIAL «MIR»

Impreso en la URSS 1974

На испанском языке

INDICE GENERAL

Parte segunda

MAGNITUDES ALEATORIAS

Parte tercera

ELEMENTOS DE ESTADISTICA MATEMATICA

INTRODUCCION

Objeto de la teoría de las probabilidades. Los sucesos (fenómenos) que observamos se pueden dividir en los tres tipos siguientes: ciertos, imposibles y aleatorios.

Se llama *cierto* el suceso que ocurrirá indefectiblemente, si se cumple un conjunto determinado de condiciones S.

Por ejemplo, si un recipiente contiene agua a la presión atmosférica normal y a la temperatura de 20°, el suceso de que el «agua del recipiente se encuentra en estado líquido» es verdadero (cierto). En este ejemplo, la presión atmosférica y la temperatura dadas del agua constituyen el conjunto de condiciones S.

Se llama *imposible* (incierto o falso) el suceso que con certeza no ocurrirá si se cumple el conjunto de condiciones S.

Por ejemplo, el suceso de que el «agua del recipiente se encuentra en estado sólido» no se producirá si se cumple el conjunto de condiciones del ejemplo anterior.

Se llama *aleatorio* el suceso que, cumpliéndose el conjunto de condiciones S, o puede ocurrir, o bien dejar de ocurrir.

Por ejemplo, si se arroja una moneda, ésta puede caer de manera que hacia arriba sea o bien cara, o bien cruz. Por eso, el suceso de que «al arrojar la moneda caiga cara» es aleatorio.

Cada suceso aleatorio (fortuito), en particular, la caída de cara, se debe a la acción de numerosas causas fortuitas (en nuestro ejemplo: la fuerza con que se arrojó la moneda, la forma de la moneda, etc.). No se puede tener en cuenta el efecto de todas estas causas en el resultado, ya que su número es muy grande y las regularidades de su acción son desconocidas. Por eso, la teoría de la probabilidad no se plantea predecir si el suceso único tendrá lugar o no, es decir, no tiene posibilidad de hacerlo.

El problema es otro si se examinan sucesos aleatorios, que es posible observar reiteradamente al producirse las mismas condiciones S, es decir, si se trata de sucesos aleatorios seme-

jantes de masas. Resulta que, un número suficientemente grande de sucesos aleatorios semejantes, independientemente de su naturaleza concreta, obedece a leyes o regularidades determinadas, precisamente, a leyes de la probabilidad. La teoría de las probabilidades se ocupa necesariamente de estas leyes.

Por tanto, *el objeto de la teoría de las probabilidades es el estudio de las leyes de la probabilidad de sucesos aleatorios semejantes en masa*.

El conocimiento de las leyes, a las que obedecen los sucesos aleatorios de masas, permite prever cómo ocurrirán estos sucesos. Por ejemplo, aunque, como se dijo antes, no se puede determinar de antemano el resultado de un solo arroje de la moneda, pero sí es posible predecir, además con pequeño error, el número de apariciones de la cara, si la moneda es arrojada un número suficientemente grande de veces. Claro está que, en este caso se supone que la moneda se arroja en idénticas condiciones.

Los métodos de la teoría de las probabilidades se utilizan ampliamente en las distintas ramas de las ciencias naturales y de la técnica: en la teoría de la fiabilidad, la teoría del servicio de masas, en física teórica, geodesia, astronomía, teoría del tiro, teoría de los errores de observaciones, teoría del mando automático, teoría general de las comunicaciones y en muchas otras ciencias teóricas y aplicadas. La teoría de las probabilidades sirve también como base de la estadística matemática y aplicada, la que, a su vez, se emplea en la planificación y organización de la producción, al analizar procesos tecnológicos, el control preventivo y de recepción de la calidad de la producción y para muchos otros fines.

En los últimos años los métodos de la teoría de las probabilidades penetran cada vez con mayor amplitud en los distintos campos de la ciencia y la técnica, contribuyendo a su progreso.

Breve información histórica. Los primeros trabajos, en los que tuvieron origen los conceptos fundamentales de la teoría de las probabilidades, se redujeron a las tentativas de crear la teoría de los juegos de azar (Cardano, Huygens, Pascal, Fermat y otros en los siglos XVI-XVII).

La etapa siguiente de desarrollo de la teoría de las probabilidades está vinculada con el nombre de Jacobo Bernoulli (1654-1705). El teorema por él demostrado, que ulteriormente tomará el nombre de «Ley de los grandes números», fue la primera base teórica de los hechos acumulados antes.

Los progresos ulteriores de la teoría de las probabilidades se deben a Moivre, Laplace, Gauss, Poisson, etc.

El nuevo período, el más fructífero, está relacionado con los nombres de P. L. Chebishev (1821-1894) y sus alumnos A. A. Markov (1856-1922) y A. M. Liapunov (1857-1918). En este período la teoría de las probabilidades se convierte en una ciencia matemática ordenada. Su desarrollo posterior se debe, en primer término, a los matemáticos rusos y soviéticos (S. N. Bernshtein, V. I. Romanovski, A. N. Kolmogorov, A. Ya. Jinchin, B. V. Gnedenko, N. V. Smirnov, etc.). Actualmente corresponde, también, a los matemáticos soviéticos el papel principal en la creación de nuevas ramas de la teoría de las probabilidades.



Parte primera

Sucesos aleatorios

Capítulo primero

CONCEPTOS FUNDAMENTALES DE LA TEORIA DE LAS PROBABILIDADES

§ 1. Experimentos y sucesos

Hemos denominado el suceso como aleatorio, si al cumplirse un conjunto determinado de condiciones S éste puede ocurrir, o dejar de ocurrir. En adelante en lugar de decir que el «conjunto de condiciones S se ha efectuado», vamos a decir que «se ha producido el experimento». De este modo, consideraremos el suceso como resultado del experimento.

Ejemplo 1. Un tirador dispara al blanco, dividido en cuatro zonas. El disparo es el experimento. El impacto en una zona determinada del blanco es el suceso.

Ejemplo 2. En una urna hay bolillas de color. De la urna tomamos al azar una bolilla. La extracción de la bolilla de la urna es el experimento. La aparición de una bolilla de un color determinado es el suceso.

§ 2. Tipos de sucesos aleatorios

Los sucesos se llaman *mutuamente excluyentes*, si la producción de uno de ellos excluye la aparición de los demás sucesos en un mismo experimento.

Ejemplo 1. De una caja de piezas se ha extraído al azar una de ellas. La aparición de la pieza standard (estándar) elimina la de la pieza no standard. Los sucesos de «aparición de la pieza standard» y de «aparición de la pieza no standard» son mutuamente excluyentes.

Ejemplo 2. Se ha arrojado una moneda. La aparición de cara excluye la aparición de cruz. Los sucesos «aparición de cara» y «aparición de cruz» se excluyen mutuamente.

Los sucesos se llaman *únicamente posibles*, si la aparición en el resultado del experimento de uno y solamente de uno de ellos es un suceso cierto.

Es evidente que, los sucesos únicamente posibles son mutuamente excluyentes.

Ejemplo 3. Se han adquirido dos billetes de lotería. Indefectiblemente ocurrirá uno y solamente uno de los siguientes sucesos: «el premio cayó en el primer billete y no cayó en el segundo», «salió premiado el segundo billete», «el premio cayó en ambos billetes», «ninguno de los billetes fue premiado». Estos sucesos son los únicos posibles.

Ejemplo 4. Un tirador disparó al blanco. Obligatoriamente ocurrirá uno de los siguientes dos sucesos: impacto o tiro perdido. Estos sucesos son los únicamente posibles.

Los sucesos se llaman *igualmente posibles*, si existe razón para suponer que ninguno de estos sucesos tiene más posibilidad que los demás.

Ejemplo 5. La aparición de cara y la aparición de cruz al arrojar una moneda son sucesos igualmente posibles. En efecto, se supone que la moneda es de una sustancia homogénea, tiene forma perfectamente cilíndrica y la acuñación no influye en la caída de uno u otro lado de la moneda.

Ejemplo 6. La aparición de uno u otro número de puntos al tirar el dado son sucesos igualmente posibles. En efecto, se supone que el dado está hecho de una sustancia homogénea, tiene forma de poliedro regular y la existencia de los puntos no influye en la caída de una u otra cara.

§ 3. Definición clásica de la probabilidad

La probabilidad es uno de los conceptos fundamentales de la teoría de las probabilidades. Existen varias definiciones de este concepto. Aquí se dará la definición llamada clásica. Más adelante (§ 6) indicaremos las deficiencias de esta definición y daremos otra definición (estadística) de probabilidad que permite superar los inconvenientes de la definición clásica.

Veamos un ejemplo. Supongamos que en una urna hay 6 bolillas idénticas, bien mezcladas; además 2 de ellas son rojas, 3 azules y 1 blanca. Evidentemente, la posibilidad de sacar al azar de la urna una bolilla de color (es decir, roja o azul) es mayor, que la posibilidad de extraer la bolilla blanca. ¿Se podrá caracterizar esta posibilidad por un número? Resulta que, se puede. Precisamente este número se llama probabilidad del suceso. Por tanto, la probabilidad es un número que caracteriza la posibilidad de que se produzca un suceso.

Planteémonos dar una estimación cuantitativa de la posibilidad de que, la bolilla extraída al azar será de color. La aparición de una bolilla de color la consideraremos como suceso A. Cada uno de los resultados posibles del experimento (éste consiste en extraer la bolilla de la urna), es decir, cada suceso, que puede ocurrir en la prueba, lo denominamos resultado elemental. Los resultados elementales los designamos por E_1, E_2, E_3, etc. En nuestro ejemplo son posibles los siguientes 6 resultados elementales (o indivisibles): E_1, apareció la bolilla blanca; E_2, E_3, apareció una bolilla roja; E_4, E_5, E_6, apareció una bolilla azul.

Se aprecia fácilmente que estos resultados son los únicamente posibles (obligatoriamente aparece una bolilla) e igualmente posibles (la bolilla se extrae al azar, las bolillas son idénticas y bien mezcladas).

Los resultados elementales, para los cuales se produce el suceso que nos interesa, los llamamos favorables a ese suceso. En nuestro ejemplo son favorables al suceso A (aparición de una bolilla de color) los 5 resultados siguientes: E_2, E_3, E_4, E_5, E_6.

La relación entre el número de resultados elementales que son favorables al suceso A y su número total se llama probabilidad del suceso A y se designa por $P(A)$. En el ejemplo considerado los resultados elementales son en total 6, y de ellos 5 son favorables al suceso A. Por lo tanto, la probabilidad de que la bolilla escogida resulte de color, es igual a $P(A) = \frac{5}{6}$.

El número hallado (probabilidad) da precisamente la estimación cuantitativa de la posibilidad de aparición de la bolilla de color, que nos propusimos hallar.

Damos ahora la definición de probabilidad.

Se llama *probabilidad de un suceso A* la relación del número de resultados que son favorables a este suceso al número total de resultados elementales únicos posibles e igualmente posibles del experimento.

De este modo, la probabilidad del suceso A se determina por la fórmula

$$P(A) = \frac{m}{n},$$

donde m es el número de resultados elementales favorables al suceso A; n es el número de todos los resultados elementa-

les posibles de la prueba. Aquí se supone que los resultados elementales son únicamente posibles e igualmente posibles.

De la definición de probabilidad se deducen sus siguientes propiedades:

1. *La probabilidad de un suceso cierto es igual a la unidad.*

En efecto, si el suceso es verdadero, cada resultado elemental de una prueba es favorable al suceso. En este caso $m = n$ y, por lo tanto,

$$P(A) = \frac{m}{n} = \frac{n}{n} = 1.$$

2. *La probabilidad de un suceso imposible es igual a cero.*

En efecto, si el suceso es imposible, ninguno de los resultados elementales de la prueba es favorable al suceso. En este caso $m = 0$ y, por lo tanto,

$$P(A) = \frac{m}{n} = \frac{0}{n} = 0.$$

3. *La probabilidad de un suceso aleatorio es un número positivo, comprendido entre cero y la unidad.*

En efecto, sólo una parte del número total de resultados elementales de la prueba es favorable al suceso aleatorio. En este caso, $0 < m < n$, vale decir, $0 < \frac{m}{n} < 1$ y, por lo tanto,

$$0 < P(A) < 1.$$

Así, la probabilidad de cualquier suceso satisface las desigualdades

$$0 \leqslant P(A) \leqslant 1.$$

Más adelante se exponen los teoremas, que simplifican considerablemente la resolución de muchos problemas. Mientras tanto damos ejemplos, para la resolución de los cuales su utiliza solamente la definición de probabilidad.

§ 4. Ejemplos de cálculo directo de probabilidades

Ejemplo 1. Al marcar el número de teléfono, el abonado se olvidó una cifra y la marcó al azar. Hallar la probabilidad de que se ha marcado la cifra necesaria.

SOLUCION. Designamos por A el suceso de haber marcado la cifra necesaria.

El abonado pudo marcar cualquiera de las 10 cifras, por eso el número total de resultados elementales posibles es 10.

Estos resultados son únicamente posibles (una de las cifras se ha marcado obligatoriamente) e igualmente posibles (cifra marcada al azar).

Sólo un resultado es favorable al suceso A (la cifra necesaria es sólo una).

La posibilidad buscada es igual a la relación entre el número de resultados, que favorecen al suceso, y el número de todos los resultados elementales

$$P(A) = \frac{1}{10}.$$

Ejemplo 2. Al marcar el número de teléfono, el abonado se olvidó las dos últimas cifras y, recordando solamente que estas cifras son diferentes, las marcó al azar. Hallar la probabilidad de que se han marcado las cifras necesarias.

SOLUCION. Designamos por B el suceso de haber marcado las dos cifras necesarias.

En total es posible marcar tantos pares de distintas cifras, como se puedan formar de diez cifras tomadas de a dos, es decir $A_{10}^2 = 10 \cdot 9 = 90$. Por consecuencia, el número total de resultados elementales posibles es igual a 90. Estos resultados son únicamente posibles e igualmente posibles. Sólo un resultado es favorable al suceso B.

La probabilidad buscada es igual a la relación del número de resultados, que son favorables al suceso, al número de todos dos resultados elementales:

$$P(B) = \frac{1}{90}.$$

Ejemplo 3. Indicar el error de la «solución» del problema: «Se han tirado dos dados. Hallar la probabilidad de que la suma de puntos aparecidos sea igual a 4 (suceso A)».

SOLUCION. En total son posibles 2 respuestas del experimento: la suma de puntos aparecidos es igual a 4, la suma de puntos aparecidos no es igual a 4. Puesto que al suceso A es favorable un resultado, y el número total de resultados es igual a dos, la probabilidad buscada es $P(A) = \frac{1}{2}$.

El error de esta solución reside en que, los resultados examinados no son igualmente posibles.

SOLUCION CORRECTA. El número total de resultados igualmente posibles del experimento es igual a $6 \cdot 6 = 36$ (cada número de puntos aparecidos en un dado se puede combinar con todos los números de puntos del otro dado). Entre

estos resultados favorecen al suceso A sólo 3 resultados: (1; 3), (3; 1), (2; 2) (entre paréntesis se indican los puntos aparecidos). Por lo tanto, la probabilidad buscada es

$$P(A)=\frac{3}{36}=\frac{1}{12}.$$

Ejemplo 4. En una partida de 10 piezas hay 7 estandarizadas. Hallar la probabilidad de que entre seis piezas tomadas al azar, justamente 4 son estandarizadas.

SOLUCIÓN. El número total de resultados elementales posibles de la prueba es igual al número de maneras, con las cuales se pueden extraer 6 piezas de 10, es decir, al número de combinaciones de 10 elementos tomados de a 6 por vez (C_{10}^6).

Calculamos el número de resultados favorables al suceso A que nos interesa, entre seis piezas tomadas al azar exactamente 4 son estandarizadas: las 4 piezas standard se pueden tomar de 7 piezas standard de C_7^4 maneras; en este caso las demás $6-4=2$ piezas deben ser no standard; 2 piezas no standard se pueden tomar de $10-7=3$ piezas no standard de C_3^2 maneras. Por lo tanto, el número de resultados favorables es igual a $C_7^4 \cdot C_3^2$. La probabilidad buscada es igual a la relación del número de resultados favorables al suceso, al número de todos los resultados elementales:

$$P(A)=\frac{C_7^4 \cdot C_3^2}{C_{10}^6}=\frac{1}{2}.$$

§ 5. Frecuencia relativa. Estabilidad de la frecuencia relativa

La frecuencia relativa, del mismo modo que la probabilidad, corresponde a los conceptos fundamentales de la teoría de las probabilidades.

Se llama *frecuencia relativa* del suceso la relación del número de experimentos o pruebas, en los cuales el suceso ha aparecido, al número total de pruebas realmente efectuadas. De esa manera, la frecuencia relativa del suceso A se determina por la fórmula

$$W(A)=\frac{m}{n},$$

donde m es el número de apariciones del suceso, n es el número total de pruebas.

Comparando las definiciones de probabilidad y de frecuencia relativa, deducimos: la definición de probabilidad

no exige que en realidad se efectúen los experimentos; en tanto que la definición de frecuencia relativa presupone que los experimentos fueron verdaderamente realizados. En otras palabras, *la probabilidad se calcula antes del experimento, mientras que la frecuencia relativa, después del experimento.*

Ejemplo 1. La Sección de control técnico descubrió 3 piezas no standard entre 80 piezas tomadas fortuitamente. La frecuencia relativa de aparición de las piezas no standard es

$$W(A)=\frac{3}{80}.$$

Ejemplo 2. Se han efectuado 24 tiros al blanco, registrándose 19 impactos. La frecuencia relativa de impacto al blanco es

$$W(A)=\frac{19}{24}.$$

Observaciones prolongadas demostraron que si las pruebas se realizan en iguales condiciones, en cada una de las cuales el número de experimentos es suficientemente grande, la frecuencia relativa exterioriza la propiedad de estabilidad. Esta propiedad consiste en que *en distintas pruebas la frecuencia relativa varía poco (tanto menos, cuanto mayor es el número de experimentos realizados), oscilando alrededor de cierto número constante.* Resulta que este número constante es la probabilidad de aparición del suceso.

Por consecuencia, si mediante la prueba se ha establecido la frecuencia relativa, el número obtenido se puede tomar como valor aproximado de la probabilidad. Más adelante se expondrá con detalle y mayor precisión la ligazón entre la frecuencia relativa y la probabilidad. A continuación ilustraremos la propiedad de estabilidad en ejemplos.

Ejemplo 3. Según datos estadísticos suecos la frecuencia relativa de nacimiento de niñas por mes en el año 1935 se caracteriza por los números siguientes (los números están dispuestos en orden de sucesión de los meses, comenzando de enero): 0,486; 0,489; 0,490; 0,471; 0,478; 0,482; 0,462; 0,484; 0,485; 0,491; 0,482; 0,473.

La frecuencia relativa oscila alrededor del número 0,482, que puede admitirse como valor aproximado de la probabilidad de nacimiento de niñas. Cabe hacer notar que los datos estadísticos de distintos países dan aproximadamente el mismo valor de la frecuencia relativa.

Ejemplo 4. Repetidamente se realizaron pruebas de arrojamiento de la moneda, en las que se contaron el número de aparición de cara. En la tabla 1 se dan los resultados de varias pruebas.

Tabla 1

Número de arrojamientos	Número de aparición de cara	Frecuencia relativa
4 040	2 048	0,5069
12 000	6 019	0,5016
24 000	12 012	0,5005

Aquí las frecuencias relativas se desvían un poco del número 0,5; además, tanto menos cuanto mayor es el número de pruebas. Por ejemplo, para 4040 pruebas la desviación es igual a 0,0069 y para 24 000 pruebas, sólo a 0,0005. Teniendo en cuenta que la probabilidad de aparición de cara al arrojar la moneda es igual a 0,5, nuevamente comprobamos que la frecuencia relativa oscila alrededor de la probabilidad.

§ 6. Insuficiencia de la definición clásica de la probabilidad. Probabilidad estadística

La definición «clásica» de la probabilidad presupone que el número de resultados elementales de la prueba es finito. En la práctica, con mucha frecuencia se encuentran pruebas, cuyo número de resultados posibles es infinito. En estos casos la definición clásica es inaplicable. Ya esta circunstancia indica la insuficiencia de la definición clásica. En realidad, el inconveniente indicado puede ser evitado mediante la respectiva generalización de la definición de probabilidad.

La parte débil de la definición clásica consiste en que muy frecuentemente no se puede imaginar el resultado de la prueba en forma de un conjunto de sucesos elementales. Aún es más difícil indicar los fundamentos que permiten considerar los sucesos elementales como equiprobables. Generalmente sobre la equiprobabilidad de los resultados elementales de la prueba se deduce de las consideraciones de simetría. Así se presenta, por ejemplo, al tirar un dado, cuando se supone que el dado tiene la forma de un poliedro (cubo)

regular. Sin embargo, los problemas, en los cuales se puede partir de las consideraciones de simetría, en la práctica se encuentran muy raramente.

Por este motivo, a la par de la definición clásica se utiliza también la definición estadística de la probabilidad, admitiendo como probabilidad del suceso la frecuencia relativa o un número próximo a ella. Por ejemplo, si como resultado de un número suficientemente grande de pruebas aparece que la frecuencia relativa es próxima al 0,4, este número se puede admitir como probabilidad estadística del suceso.

Problemas

1. En una caja hay 50 piezas idénticas, 5 de las cuales están pintadas. Extraemos una pieza al azar. Hallar la probabilidad de que la pieza extraída resulte pintada.

Respuesta $p = 0{,}1$.

2. Tirado un dado, hallar la probabilidad de que aparezca un número par de puntos.

Respuesta $p = 0{,}5$.

3. Los participantes de un sorteo sacan de una caja fichas numeradas desde 1 hasta 100. Hallar la probabilidad de que la primera ficha extraída al azar no contiene la cifra 5.

Respuesta $p = 0{,}81$.

4. En una bolsa hay 5 cubos idénticos. En todas las caras de cada cubo está escrita una de las letras siguientes: o, p, r, s, t. Hallar la probabilidad de que en los cubos extraídos de a uno por vez y dispuestos «en una línea» se pueda leer la palabra «sport».

Respuesta $p = \frac{1}{120}$.

5. En cada una de las seis tarjetas idénticas está impresa una de las letras siguientes: a, t, m, r, s, e. Las tarjetas están bien mezcladas. Hallar la probabilidad de que en cuatro tarjetas extraídas de a una por vez y dispuestas «en una línea», se pueda leer la palabra «tres»

Respuesta $p = \frac{1}{A_6^4} = \frac{1}{360}$.

6. Un cubo, cuyas caras están pintadas, se ha dividido en mil cubos más pequeños de igual dimensión y se han mezclado cuidadosamente. Hallar la probabilidad de que el cubo tomado al azar tendrá las caras pintadas: a) una; b) dos; c) tres.

Respuestas a) 0,384; b) 0,096; c) 0,008.

7. De un juego completo de 28 fichas de dominó bien mezcladas se ha extraído al azar una ficha. Hallar la probabilidad de que la segun-

da ficha tomada al azar se pueda juntar a la primera, si ésta: a) resulta el doble; b) no es el doble.

Respuestas a) $p=\frac{2}{9}$; b) $\frac{4}{9}$.

8. Sobre el eje común de una cerradura hay cinco discos, cada uno de los cuales está dividido en seis sectores con distintas letras incriptas en ellos. La cerradura se abre sólo cuando cada disco ocupa una posición determinada con respecto al cuerpo de la cerradura. Hallar la probabilidad de que, al instalar arbitrariamente los discos la cerradura pueda ser abierta.

Respuesta $p=\frac{1}{6^5}$.

9. Ocho libros diferentes se ponen al azar en un estante. Hallar la probabilidad de que dos libros determinados resulten puestos juntos.

Respuesta $p=\frac{7\cdot 2!\cdot 6!}{8!}=\frac{1}{4}$.

10. Una pequeña biblioteca está formada de diez libros diferentes; además, cinco libros cuestan 4 rublos cada uno, tres libros, a un rublo cada uno y dos libros, a 3 rublos c/u. Hallar la probabilidad de que dos libros tomados al azar cuesten 5 rublos.

Respuesta $p=\frac{C_5^1\cdot C_3^1+C_2^1\cdot C_3^1}{C_{10}^2}=\frac{7}{15}$.

11. En la partida de 100 piezas la sección de control técnico descubrió 5 piezas no estandarizadas. ¿Cuál es la frecuencia relativa de aparición de las piezas no estandarizadas?

Respuesta $w = 0{,}05$.

12. Al hacer fuego con un fusil, la frecuencia relativa de impacto al blanco resultó igual a 0,85. Hallar el número de impactos, si en total se efectuaron 120 disparos.

Respuesta 102 impactos.

Capítulo segundo

TEOREMA DE LA ADICION DE PROBABILIDADES

§ 1. Teorema de la adición de probabilidades de sucesos mutuamente excluyentes

Se llama *suma $A+B$ de dos sucesos A* y B el suceso, compuesto de la aparición del suceso A o del suceso B, o ambos sucesos.

Por ejemplo, si de un arma se han hecho dos disparos, siendo A el impacto en el primer disparo y B, el impacto en

el segundo disparo, tendremos que $A + B$ es el impacto en el primer disparo, o en el segundo, o bien en ambos disparos.

En particular, si dos sucesos A y B se excluyen mutuamente, $A + B$ es un suceso compuesto de la aparición de uno de esos sucesos, indiferentemente cual de ellos sea.

Se llama *suma de varios sucesos* el suceso compuesto de la aparición aunque sea de uno de estos sucesos.

Por ejemplo, el suceso $A + B + C$ está compuesto de la aparición de uno de los sucesos siguientes: A, B, C; A y B; A y C; B y C; A y B y C.

Supongamos que los sucesos A y B se excluyen mutuamente; además, están dadas las probabilidades de estos sucesos. ¿Cómo hallar la probabilidad de que ocurrirá el suceso A, o el suceso B? La respuesta a esta pregunta la da el teorema de la adición.

Teorema. *La probabilidad de que ocurra uno de los dos sucesos que se excluyen recíprocamente, indiferentemente cual de ellos, es igual a la suma de las probabilidades de esos sucesos:*

$$P(A + B) = P(A) + P(B).$$

DEMOSTRACION. Designamos por:

n, el número total de resultados elementales posibles del experimento;

m_1, el número de resultados, que favorecen al suceso A;

m_2, el número de resultados, que favorecen al suceso B.

El número de resultados elementales que favorecen a la producción del suceso A, o del suceso B, es igual a $m_1 + m_2$.

Por lo tanto,

$$P(A+B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n}.$$

Teniendo en cuenta que $\frac{m_1}{n} = P(A)$ y $\frac{m_2}{n} = P(B)$, obtenemos finalmente

$$P(A + B) = P(A) + P(B).$$

Corolario. *La probabilidad de que ocurra uno de varios sucesos que se excluyen recíprocamente a pares, indiferentemente cuál de ellos, es igual a la suma de las probabilidades de estos sucesos:*

$$P(A_1 + A_2 \dots + A_n) = P(A_1) + P(A_2) + \dots$$
$$\dots + P(A_n).$$

DEMOSTRACION. Examinemos tres sucesos A, B y C. Puesto que los sucesos considerados se excluyen mutuamente

a pares, la aparición de uno de los tres sucesos A, B y C, equivale a la producción de uno de los dos sucesos $A + B$ y C, por eso, en virtud del teorema indicado,

$$P(A + B + C) = P[(A + B) + C] =$$
$$= P(A + B) + P(C) = P(A) + P(B) + P(C).$$

Para un número arbitrario de sucesos mutuamente excluyentes a pares la demostración se realiza por el método de inducción matemática.

Ejemplo 1. En una urna hay 30 bolillas: 10 rojas, 5 azules y 15 blancas. Hallar la probabilidad de aparición de una bolilla de color.

SOLUCION. La aparición de una bolilla de color significa la aparición de una bolilla roja o azul.

La probabilidad de aparición de una bolilla roja (suceso A) es

$$P(A) = \frac{10}{30} = \frac{1}{3}.$$

La probabilidad de aparición de una bolilla azul (suceso B) es

$$P(B) = \frac{5}{30} = \frac{1}{6}.$$

Los sucesos A y B se excluyen recíprocamente (la aparición de la bolilla de un color excluye la aparición de una bolilla de otro color), por eso el teorema de la suma es aplicable. La probabilidad buscada es

$$P(A+B) = P(A) + P(B) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

Ejemplo 2. Un tirador tira al blanco, dividido en tres zonas. La probabilidad de impacto en la primera zona es igual a 0,45, en la segunda, 0,35. Hallar la probabilidad de que el tirador con un disparo hace impacto o bien en la primera zona, o bien en la segunda.

SOLUCION. El suceso A, ocurre cuando «el tirador hace impacto en la primera zona», y B, cuando «el tirador hace impacto en la segunda zona» se excluyen mutuamente (el

impacto en una zona excluye el impacto en la otra), por eso es aplicable el teorema de la adición.

La probabilidad buscada es

$$P(A + B) = P(A) + P(B) = 0{,}45 + 0{,}35 = 0{,}80.$$

§ 2. Grupo completo de sucesos

Se llama *grupo completo* el conjunto de sucesos únicamente posibles del experimento.

Ejemplo 1. Un tirador hace 2 disparos al blanco. Los sucesos A_1 (un impacto), A_2 (2 impactos) y A_3 (tiro fallado) forman un grupo completo.

Teorema. *La suma de las probabilidades de los sucesos A_1, A_2, . . ., A_n, que forman un grupo completo, es igual a la unidad:*

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1.$$

DEMOSTRACION. Ya que la aparición de uno de los sucesos del grupo completo es cierta, y la probabilidad de un suceso cierto es igual a la unidad, tendremos que

$$P(A_1 + A_2 + \ldots + A_n) = 1. \qquad (*)$$

Dos sucesos cualesquiera del grupo total se excluyen mutuamente, por eso se puede aplicar el teorema de la adición:

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \\ + \ldots + P(A_n). \qquad (**)$$

Confrontando (*) y (**), obtenemos

$$P(A_1) + P(A_2) + \ldots + P(A_n) = 1.$$

Ejemplo 2. El centro de consulta de un instituto recibe paquetes con trabajos de control desde las ciudades A, B y C. La probabilidad de recibir un paquete de la ciudad A es igual a 0,7, de la ciudad B, igual a 0,2. Hallar la probabilidad de que el paquete siguiente se recibirá de la ciudad C.

SOLUCION. Los sucesos «un paquete ha sido recibido de la ciudad A», «un paquete ha sido recibido de la ciudad B» y «un paquete ha sido recibido de la ciudad C» forman un grupo completo, por eso, la suma de las probabilidades de

estos sucesos es igual a la unidad:

$$0{,}7 + 0{,}2 + p = 1.$$

De donde, la probabilidad buscada es

$$p = 1 - 0{,}9 = 0{,}1.$$

§ 3. Sucesos opuestos

Dos sucesos únicamente posibles que forman un grupo completo se llaman *opuestos*. Si uno de los dos sucesos opuestos se designa por A, el otro se admite en designar por $\bar{A}$.

Ejemplo 1. El impacto y el fallo en el tiro al blanco son sucesos opuestos. Si A es el impacto, $\bar{A}$ es el tiro fallado.

Ejemplo 2. De una caja se ha tomado al azar una pieza. Los sucesos «apareció una pieza standard» y «apareció una pieza no standard» son opuestos.

Teorema. *La suma de las probabilidades de sucesos opuestos es igual a la unidad:*

$$P(A) + P(\bar{A}) = 1.$$

DEMOSTRACIÓN. Los sucesos opuestos forman un grupo completo, y la suma de las probabilidades de los sucesos que forman un grupo completo es igual a la unidad (§ 2).

$$p + q = 1.$$

Nota. 1. Si la probabilidad de uno de los dos sucesos opuestos se designa por p, la probabilidad del otro suceso se designa por q. De este modo, en virtud del teorema anterior

$$p + q = 1.$$

Ejemplo 3. La probabilidad de que el día será lluvioso es $p = 0{,}7$. ¿Cuál es la probabilidad de que el día sea claro?

SOLUCIÓN. Los sucesos «día lluvioso» y «día claro» son opuestos, por eso la probabilidad buscada es

$$q = 1 - p = 1 - 0{,}7 = 0{,}3.$$

Nota 2. Al resolver el problema de hallar la probabilidad del suceso A, de ordinario conviene al principio calcular la probabilidad del suceso $\bar{A}$, y después hallar la probabilidad buscada por la fórmula:

$$P(A) = 1 - P(\bar{A}).$$

Ejemplo 4. En una caja hay n piezas, de las cuales m son standard. Hallar la probabilidad de que entre k piezas extraídas al azar hay aunque sea una standard.

SOLUCION. Los sucesos «entre las piezas extraídas hay aunque sea una standard» y «entre las piezas extraídas no hay ni una standard» son opuestos. Designamos el primer suceso por A, y el segundo por $\overline{A}$.

Evidentemente

$$P(A)=1-P(\overline{A}).$$

Hallamos $P(\overline{A})$. El número total de maneras, mediante las cuales se pueden extraer k piezas de n piezas, es igual a C_n^k. El número de piezas no standard es igual a $n-m$; de este número de piezas se pueden de C_{n-m}^k maneras extraer k piezas no standard. Por eso, la probabilidad de que entre las k piezas extraídas no hay ni una standard, es igual a $P(\overline{A})=\frac{C_{n-m}^k}{C_n^k}$.

La probabilidad buscada es

$$P(A)=1-P(\overline{A})=1-\frac{C_{n-m}^k}{C_n^k}.$$

§ 4. Principio de imposibilidad práctica de sucesos poco probables

Al resolver muchos problemas prácticos se tropiezan con sucesos, cuya probabilidad es muy poca, es decir, próxima a cero. ¿Podría considerarse que el suceso poco probable A no ocurrirá en una tentativa única? No se puede hacer tal deducción, ya que no se excluye, aunque es poco probable, que el suceso A ocurrirá.

Aparentemente, no es posible predecir si ocurrirá o no un suceso poco probable en una tentativa (experimento). Sin embargo, la larga experiencia muestra que el suceso poco probable en una tentativa única en la mayoría de los casos no ocurre. Basándose en este hecho se admite el siguiente «principio de imposibilidad práctica de los sucesos poco probables»: *si un suceso aleatorio tiene muy poca probabilidad de que ocurra, prácticamente se puede considerar que en una tentativa única este suceso no ocurrirá.*

Naturalmente surge la pregunta: ¿cuán poca debe ser la probabilidad de un suceso para poder considerar que es im-

posible que ocurra en un experimento? A esta pregunta no se puede responder simplemente. Las respuestas para los distintos problemas serán diferentes. Por ejemplo, si la probabilidad de que el paracaídas no se abre al saltar, es igual a 0,01, sería inadmisible utilizar semejante paracaídas. Si la probabilidad de que un tren de largo recorrido llegue con retardo es igual a 0,01, prácticamente se puede tener la certidumbre de que el tren arribará a tiempo.

La probabilidad suficientemente pequeña, para la cual (en un problema dado) el suceso puede considerarse prácticamente imposible, se llama *nivel de significación*. De ordinario, en la práctica se utilizan niveles de significación comprendidos entre 0,01 y 0,05. El nivel de significación, igual a 0,01, se llama de uno por ciento; el nivel de significación, igual a 0,02, se llama de dos por ciento, etc.

Cabe señalar que el principio aquí examinado permite predecir no sólo los sucesos, que tienen poca probabilidad, sino también los sucesos, cuya probabilidad es próxima a la unidad. En efecto, si el suceso A tiene una probabilidad próxima a cero, la probabilidad del suceso opuesto $\overline{A}$ es próxima a la unidad. Por otro lado, la no aparición del suceso A significa la aparición del suceso opuesto $\overline{A}$. De este modo, del principio de imposibilidad de los sucesos poco probables se deduce el siguiente corolario importante para las aplicaciones: *si un suceso aleatorio tiene probabilidad muy próxima a la unidad, prácticamente se puede considerar que este suceso ocurrirá en un experimento*. Está claro que también aquí la respuesta a la pregunta, cuál probabilidad hay que considerar próxima a la unidad, depende de la naturaleza del problema.

Problemas

1. A cada 10 000 billetes de lotería se juegan 150 premios en objetos y 50 premios en dinero. ¿Qué propabilidad tiene de ganar, indiferentemente dinero u objeto, el poseedor de un billete de lotería?

Respuesta $p = 0,02$.

2. La probabilidad de que un tirador en un disparo marque 10 puntos es igual a 0,1; la probabilidad de marcar 9 puntos es igual a 0,3; la probabilidad de marcar 8 o menos puntos es igual a 0,6. ¿Cuál es la probabilidad de que en un disparo el tirador marque no menos de 9 puntos?

Respuesta $p = 0,4$.

3. En un lote de 10 piezas 8 son standard. Hallar la probabilidad de que entre 2 piezas tomadas al azar aunque sea una es standard.

Respuesta $p=\frac{44}{45}$.

4. En una caja hay 10 piezas, entre las cuales 2 son no standard. ¿Cual es la probabilidad de que entre 6 piezas escogidas al azar no más de una resulte una pieza no standard?

Respuesta $p=\frac{2}{3}$.

Advertencia. Si A significa que no hay ninguna pieza no standard y B, hay una pieza no standard, tendremos que

$$P(A+B)=P(A)+P(B)=\frac{C_8^6}{C_{10}^6}+\frac{C_2^1\cdot C_8^5}{C_{10}^6}.$$

5. Los sucesos A, B, C y D forman un sistema completo. Las probabilidades de los sucesos son: $P(A)=0{,}1$; $P(B)=0{,}4$; $P(C)=0{,}3$. ¿A qué es igual la probabilidad del suceso D?

Respuesta $P(D)=0{,}2$.

6. Según datos estadísticos de un taller de reparaciones, en un promedio de 20 paradas de un torno se encuentran: 10 para cambiar la cuchilla; 3 debido al mal estado de la transmisión; 2 por el suministro a destiempo de la pieza bruta. Las demás paradas ocurren por otros motivos. ¿Cual es la probabilidad de parada del torno por otros motivos?

Respuesta $p=0{,}25$.

Capítulo tercero

TEOREMA DEL PRODUCTO DE PROBABILIDADES

§ 1. Sucesos independientes y dependientes

Dos sucesos se llaman *independientes*, cuando la probabilidad de uno de ellos no depende de la aparición o no del otro.

Ejemplo 1. Una moneda se arroja 2 veces. La probabilidad de que aparezca cara en la primera prueba (suceso A) no depende de la aparición o no aparición de cara en la segunda prueba (suceso B). A su vez, la probabilidad de que en la segunda prueba caiga cara no depende del resultado de la primera prueba. En consecuencia, los sucesos A y B son independientes.

Ejemplo 2. En una urna hay 5 bolillas blancas y 3 negras. Se extrae una bolilla al azar. Evidentemente, la probabilidad de que aparezca una bolilla blanca (suceso A) es igual a $\frac{5}{8}$. La bolilla extraída se vuelve a tirar en la urna y se repite la prueba. La probabilidad de que aparezca una bolilla blanca en la segunda tentativa (suceso B), como antes, es igual a $\frac{5}{8}$ y no depende del resultado de la primera prueba. A su vez, la probabilidad de extraer una bolilla blanca en la primera tentativa no depende del resultado de la segunda tentativa. Por tanto, los sucesos A y B son independientes.

Varios sucesos se denominan *independientes por parejas*, si cada dos de ellos son independientes.

Ejemplo 3. Una moneda se arroja 3 veces. Supongamos que A, B, y C son los sucesos, que constituyen la aparición de cara respectivamente en la primera, segunda y tercera pruebas. Está claro que, cada par de sucesos considerados (es decir, A y B, A y C, B y C) son independientes. De este modo, los sucesos A, B y C son independientes por parejas (de dos en dos).

Dos sucesos se llaman *dependientes*, cuando la probabilidad de que ocurra uno de ellos depende de que ocurra o no el otro suceso.

Ejemplo 4. En una caja hay 100 piezas: 80 standard y 20 no standard. Se toma al azar una pieza, sin volverla a colocar en la caja. Si apareció una pieza standard (suceso A), la probabilidad de extraer una pieza standard en la segunda tentativa (suceso B) es $P(B)=\frac{79}{99}$; si en la primera tentativa se extrajo una pieza no standard, la probabilidad $P(B) = \frac{80}{99}$.

Por lo tanto, la probabilidad de que ocurra el suceso B depende de que ocurra o no el suceso A. Los sucesos A y B son dependientes.

§ 2. Teorema del producto de probabilidades de sucesos independientes

Se llama *producto de dos sucesos* A y B el suceso AB, constituido por la aparición simultánea de estos sucesos.

Por ejemplo, si una caja contiene piezas producidas en las fábricas N° 1 y N° 2, y A es la aparición de una pieza

standard, B es una pieza producida en la fábrica N° 1, tendremos que AB es la aparición de una pieza standard de la fábrica N° 1.

Se llama *producto de varios sucesos* el suceso compuesto de la aparición simultánea de todos estos sucesos. Por ejemplo, el suceso ABC está compuesto de los sucesos simultáneos A, B y C.

Supongamos que los sucesos A y B son independientes; además, se conocen las probabilidades de estos sucesos. ¿Cómo hallar la probabilidad de simultaneidad de los sucesos A y B? La respuesta a esta pregunta la da el teorema del producto.

Teorema. *La probabilidad de que aparezcan simultáneamente dos sucesos independientes es igual al producto de las probabilidades de estos sucesos:*

$$P(AB) = P(A) \cdot P(B).$$

DEMOSTRACION. Introducimos las designaciones:

n, número de resultados elementales posibles de la prueba, en los cuales el suceso A puede ocurrir o no;

n_1, número de resultados favorables al suceso A $(n_1 \leqslant n)$;

m, número de resultados elementales posibles de la prueba, en los cuales el suceso B puede ocurrir o no;

m_1, número de resultados favorables al suceso B $(m_1 \leqslant m)$.

El número total de resultados elementales posibles del experimento (en los que ocurren A y B, o bien A y $\overline{B}$, o bien $\overline{A}$ y B, o bien $\overline{A}$ y $\overline{B}$) es igual a mn. En efecto, cada uno de los n resultados, en los que el suceso A ocurre o no, puede combinarse con cada uno de los m resultados, en los que el suceso B aparece o no.

De este número $n_1 m_1$ resultados son favorables a la simultaneidad de los sucesos A y B. En realidad, cada uno de los n_1 resultados favorables al suceso A, puede combinarse con cada uno de los m_1 resultados favorables al suceso B.

La probabilidad de que los sucesos A y B ocurran simultáneamente es

$$P(AB) = \frac{n_1 m_1}{nm} = \frac{n_1}{n} \cdot \frac{m_1}{m}.$$

Tomando en consideración que $\frac{n_1}{n} = P(A)$ y $\frac{m_1}{m} = P(B)$, finalmente obtenemos:

$$P(AB) = P(A) \cdot P(B).$$

Para generalizar el teorema del producto a varios sucesos, introducimos el concepto de independencia de los sucesos en el conjunto.

Varios sucesos se llaman *independientes en el conjunto,* si cada uno de ellos y cualquier combinación de los demás sucesos (que contiene todos los restantes sucesos, o parte de ellos) son sucesos independientes. Por ejemplo, si los sucesos A_1, A_2 y A_3 son independientes en el conjunto, tendremos que son independientes los sucesos: A_1, y A_2, A_1 y A_3, A_2 y A_3, A_1A_2 y A_3, A_1A_3 y A_2, A_2A_3 y A_1.

Cabe hacer notar que, si varios sucesos son independientes de dos en dos, eso aún no significa su independencia en el conjunto. En este sentido, el requisito de independencia de los sucesos en el conjunto es mayor que el requisito de su independencia por pareja.

Aclaremos lo dicho con un ejemplo. Supongamos que en una urna hay 4 bolillas pintadas: 1 de color rojo (A), 1 de color azul (B), 1 de color negro (C) y 1 de estos tres colores (ABC). ¿A qué es igual la probabilidad $P(A)$ de que la bolilla escogida de la urna es de color rojo? Puesto que de las cuatro bolillas dos tienen color rojo, tendremos que $P(A) = \frac{2}{4} = \frac{1}{2}$. Con análogo razonamiento, hallamos: $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{2}$.

Supongamos ahora que la bolilla escogida tiene color azul, es decir, que el suceso B ya ocurrió. ¿Cambia la probabilidad de que la bolilla escogida tiene color rojo, es decir, existe la probabilidad del suceso A? De dos bolillas, que tienen color azul, una bolilla tiene también color rojo, por eso la probabilidad del suceso A, como antes, es igual a $\frac{1}{2}$. Por consecuencia, los sucesos A y B son independientes.

Razonando de manera análoga, deducimos que los sucesos A y C, B y C son independientes. De este modo, los sucesos, A, B y C son independientes de dos en dos.

¿Serán independientes estos sucesos en el conjunto? Resulta que no. En efecto, supongamos que la bolilla extraída tiene dos colores, por ejemplo, azul y negro. ¿Cuál es la probabilidad de que esta bolilla tenga también color rojo? Puesto que sólo una bolilla está pintada con los tres colores, la bolilla extraída tiene también color rojo. En consecuencia, suponiendo que los sucesos B y C han ocurrido, llegamos a la

conclusión que el suceso A ocurrirá con certeza. Por lo tanto, este suceso es cierto y su probabilidad, igual a la unidad $\left(\text{y no a } \frac{1}{2}\right)$.

Así, los sucesos independientes de dos en dos A, B y C son dependientes en el conjunto.

A continuación damos el corolario del teorema del producto.

Corolario. *La probabilidad de que ocurran simultáneamente varios sucesos, independientes en el conjunto, es igual al producto de las probabilidades de estos sucesos:*

$$P(A_1A_2\ldots A_n)=P(A_1)\cdot P(A_2)\ldots P(A_n).$$

DEMOSTRACION. Examinemos los tres sucesos A, B y C. La simultaneidad de los sucesos A, B y C es equivalente a la simultaneidad de los sucesos AB y C, por eso,

$$P(ABC)=P(AB\cdot C).$$

Dado que los sucesos A, B y C son independientes en el conjunto, en particular son independientes los sucesos AB y C, así como A y B. Por el teorema del producto para dos sucesos independientes tendremos:

$$P(AB\cdot C)=P(AB)\cdot P(C)$$

y

$$P(AB)=P(A)\cdot P(B).$$

De este modo, finalmente obtenemos

$$P(ABC)=P(A)\cdot P(B)\cdot P(C).$$

Para n arbitrario la demostración se realiza por el método de inducción matemática.

Nota. Si los sucesos A_1, A_2, ..., A_n son independientes en el conjunto, los sucesos opuestos a ellos $\overline{A}_1$, $\overline{A}_2$, ..., $\overline{A}_n$ también son independientes en el conjunto.

Ejemplo 1. Hallar la probabilidad de que aparezca simultáneamente la cara al arrojar a la vez dos monedas.

SOLUCION. La probabilidad de aparecer la cara de la primera moneda (suceso A) es

$$P(A)=\frac{1}{2}.$$

La probabilidad de que aparezca la cara de la segunda moneda (suceso B) es

$$P(B)=\frac{1}{2}.$$

Ya que los sucesos A y B son independientes, por el teorema del producto la probabilidad buscada es igual a

$$P(AB)=P(A)\cdot P(B)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}.$$

Ejemplo 2. Tengamos 3 cajas que contienen 10 piezas cada una. En la primera caja 8, en la segunda 7 y en la tercera 9 son piezas standard. De cada caja se extrae al azar una pieza. Hallar la probabilidad de que las tres piezas extraídas sean standards.

SOLUCION. La probabilidad de que de la primera caja se extrajo una pieza standard (suceso A) es

$$P(A)=\frac{8}{10}=0{,}8.$$

La probabilidad de que de la segunda caja se ha extraído una pieza standard (suceso B) es

$$P(B)=\frac{7}{10}=0{,}7.$$

La probabilidad de que de la tercera caja se haya extraído una pieza standard (suceso C) es

$$P(C)=\frac{9}{10}=0{,}9.$$

Puesto que los sucesos A, B y C son independientes en el conjunto, la probabilidad buscada (por el teorema del producto) es igual a

$$P(ABC)=P(A)\cdot P(B)\cdot P(C)=0{,}8\cdot 0{,}7\cdot 0{,}9=0{,}504.$$

Damos un ejemplo de aplicación conjunta de los teoremas de adición y multiplicación.

Ejemplo 3. La probabilidad de aparición de cada uno de los tres sucesos independientes A_1, A_2 y A_3 es respectivamente igual a p_1, p_2 y p_3. Hallar la probabilidad de que aparezca sólo uno de estos sucesos.

SOLUCION. Cabe hacer notar, por ejemplo, que la aparición solamente del primer suceso A_1 equivale a la aparición

del suceso $A_1\bar{A}_2\bar{A}_3$ (apareció el primer suceso y no aparecieron el segundo y tercer sucesos).

Designemos por:

B_1, apareció sólo el suceso A_1, es decir, $B_1 = A_1\bar{A}_2\bar{A}_3$;
B_2, apareció sólo el suceso A_2, es decir, $B_2 = A_2\bar{A}_1\bar{A}_3$;
B_3, apareció sólo el suceso A_3, es decir, $B_3 = A_3\bar{A}_1\bar{A}_2$.

De este modo, para hallar la probabilidad de que aparezca solamente uno de los sucesos A_1, A_2, A_3, vamos a buscar la probabilidad $P(B_1 + B_2 + B_3)$ de que aparezca uno, indiferentemente cual de los sucesos B_1, B_2, B_3.

Dado que los sucesos B_1, B_2, B_3 son mutuamente excluyentes, es aplicable el teorema de la suma:

$$P(B_1 + B_2 + B_3) = P(B_1) + P(B_2) + P(B_3). \qquad (*)$$

Queda hallar las probabilidades de cada uno de los sucesos B_1, B_2, B_3.

Los sucesos A_1, A_2 y A_3 son independientes, por lo tanto, son independientes los sucesos A_1, $\bar{A}_2$ y $\bar{A}_3$, por eso a ellos se puede aplicar el teorema del producto:

$$P(B_1) = P(A_1\bar{A}_2\bar{A}_3) = P(A_1)\,P(\bar{A}_2)\,P(\bar{A}_3) = p_1q_2q_3.$$

Análogamente:

$$P(B_2) = P(A_2\bar{A}_1\bar{A}_3) = P(A_2)\,P(\bar{A}_1)\,P(\bar{A}_3) = p_2q_1q_3;$$

$$P(B_3) = P(A_3\bar{A}_1\bar{A}_2) = P(A_3)\,P(\bar{A}_1)\,P(\bar{A}_2) = p_3q_1q_2.$$

Sustituyendo estas probabilidades en (*), hallamos la probabilidad buscada de que aparezca sólo uno de los sucesos A_1, A_2, A_3:

$$P(B_1 + B_2 + B_3) = p_1q_2q_3 + p_2q_1q_3 + p_3q_1q_2.$$

§ 3. Probabilidad de aparición aunque sea de un suceso

Supongamos que como resultado de la prueba pueden ocurrir n sucesos independientes en el conjunto, o bien algunos de ellos (en particular, sólo uno o ninguno); además, las probabilidades de que ocurran cada uno de los sucesos son conocidas. ¿Cómo hallar la probabilidad de que ocurrirá aunque sea uno de estos sucesos? Por ejemplo, si como resultado de la prueba pueden ocurrir tres sucesos, la aparición aunque sea de uno de estos sucesos significa la aparición de

uno, dos o tres sucesos. La respuesta a esta pregunta está dada por el siguiente teorema.

Teorema. *La probabilidad de que ocurra aunque sea uno de los sucesos $A_1, A_2, \ldots, A_n$, independientes en el conjunto, es igual a la diferencia entre la unidad y el producto de las probabilidades de los sucesos opuestos $\overline{A}_1\overline{A}_2 \ldots \overline{A}_n$:*

$$P(A) = 1 - q_1q_2 \ldots q_n. \qquad (*)$$

DEMOSTRACION. Designamos por A el suceso constituido por la aparición aunque sea de uno de los sucesos A_1, $A_2, \ldots, A_n$. Los sucesos A y $\overline{A}_1\overline{A}_2 \ldots \overline{A}_n$ (no ha ocurrido ninguno de los sucesos) son opuestos, por lo tanto, la suma de sus probabilidades es igual a la unidad:

$$P(A) + P(\overline{A}_1\overline{A}_2 \ldots \overline{A}_n) = 1.$$

De aquí, utilizando el teorema del producto, obtenemos: $P(A) = 1 - P(\overline{A}_1\overline{A}_2 \ldots \overline{A}_n) = 1 - P(\overline{A}_1)\cdot P(\overline{A}_2) \ldots$ $\ldots P(\overline{A}_n)$, o bien

$$P(A) = 1 - q_1q_2 \ldots q_n.$$

CASO PARTICULAR. *Si los sucesos $A_1, A_2, \ldots, A_n$ tienen idéntica probabilidad, igual a p, la probabilidad de que ocurra por lo menos uno de estos sucesos es*

$$P(A) = 1 - q^n. \qquad (**)$$

Ejemplo 1. Las probabilidades de impacto en el blanco al disparar de tres armas son: $p_1 = 0{,}8$; $p_2 = 0{,}7$; $p_3 = 0{,}9$. Hallar la probabilidad de por lo menos un impacto (suceso A) al disparar simultáneamente de todas las armas.

SOLUCION. La probabilidad de impacto en el blanco con cada arma no depende de los resultados del disparo desde otras armas, por eso, los sucesos considerados A_1 (impacto de la primera arma), A_2 (impacto de la segunda arma) y A_3 (impacto de la tercera arma) son independientes en el conjunto.

Las probabilidades de los sucesos opuestos a los sucesos A_1, A_2 y A_3 (es decir, las probabilidades de fallos), son respectivamente iguales a:

$$q_1 = 1 - p_1 = 1 - 0{,}8 = 0{,}2;$$

$$q_2 = 1 - p_2 = 1 - 0{,}7 = 0{,}3;$$

$$q_3 = 1 - p_3 = 1 - 0{,}9 = 0{,}1.$$

La probabilidad buscada es

$$P(A) = 1 - q_1 q_2 q_3 = 1 - 0{,}2 \cdot 0{,}3 \cdot 0{,}1 = 0{,}994.$$

Ejemplo 2. En una imprenta hay 4 máquinas planas. La probabilidad de que cada máquina trabaja en el instante dado, es igual a 0,9. Hallar la probabilidad de que en el instante dado trabaje por lo menos una máquina (suceso A).

SOLUCION. Puesto que los sucesos «una máquina trabaja» y «una máquina no trabaja» (en el instante dado) son opuestos, la suma de sus probabilidades es igual a la unidad:

$$p + q = 1.$$

De donde la probabilidad de que en el instante dado la máquina no trabaja, es igual a

$$q = 1 - p = 1 - 0{,}9 = 0{,}1.$$

La probabilidad buscada es

$$P(A) = 1 - q^4 = 1 - 0{,}1^4 = 0{,}9999.$$

Puesto que la probabilidad obtenida es muy próxima a la unidad, basándose en el corolario del principio de imposibilidad práctica de los sucesos poco probables, podemos deducir que en el instante dado trabaja por lo menos una de las máquinas.

Ejemplo 3. La probabilidad de que en un disparo un tirador haga impacto en el blanco es igual a 0,4. ¿Cuántas veces debe disparar el tirador para que con una probabilidad no menor de 0,9 haga blanco por lo menos una vez?

SOLUCION. Designemos por A el suceso: en n disparos el tirador hace blanco por lo menos una vez.

Los sucesos constituidos por el impacto en el blanco en el primero, segundo, etc. disparos, son independientes en el conjunto, por lo cual es aplicable la fórmula (**)

$$P(A) = 1 - q^n.$$

Teniendo en cuenta que por la condición $P(A) \geqslant 0{,}9$, $p = 0{,}4$ (por lo tanto, $q = 1 - 0{,}4 = 0{,}6$), obtenemos:

$$1 - 0{,}6^n \geqslant 0{,}9.$$

De donde

$$0{,}6^n \leqslant 0{,}1.$$

Tomando el logaritmo decimal de esta desigualdad:

$$n \lg 0,6 \leqslant \lg 0,1.$$

De donde, tomando en consideración que $\lg 0,6 < 0$, tendremos

$$n \geqslant \frac{\lg 0,1}{\lg 0,6} = \frac{-1}{1,7782} = \frac{-1}{-0,2218} = 4,5.$$

Así, $n \geqslant 5$, es decir, el tirador ha de efectuar no menos de 5 disparos.

Ejemplo 4. La probabilidad de que el suceso ocurra por lo menos una vez en tres pruebas independientes en el conjunto es igual a 0,936. Hallar la probabilidad de que ocurra el suceso en una prueba (se supone que en todas las pruebas la probabilidad de que ocurra el suceso es idéntica).

SOLUCION. Puesto que los sucesos considerados son independientes en el conjunto, donde es aplicable la fórmula (**)

$$P(A) = 1 - q^n.$$

Por la condición $P(A) = 0,936$; $n = 3$. Por lo tanto,

$$0,936 = 1 - q^3$$

o bien

$$q^3 = 1 - 0,936 = 0,064.$$

De aquí

$$q = \sqrt[3]{0,064} = 0,4.$$

La probabilidad buscada es

$$p = 1 - q = 1 - 0,4 = 0,6.$$

§ 4. Probabilidad condicional

Supongamos que los sucesos A y B son dependientes. De la definición de sucesos dependientes se deduce que la probabilidad de uno de los sucesos depende de la aparición o no del otro. Por eso, si nos interesa la probabilidad, por ejemplo, del suceso B, es importante saber si se realizó el suceso A.

Se llama *probabilidad condicional* $P_A(B)$ la probabilidad del suceso B calculado suponiendo que el suceso A ha ocurrido ya.

Ejemplo. Una urna contiene 3 bolillas blancas y 3 negras. De la urna se extraen dos veces al azar de a una bolilla

por vez sin restituirlas en la urna. Hallar la probabilidad de que aparezca una bolilla blanca en la segunda prueba (suceso B), si en la primera prueba se extrajo una bolilla negra (suceso A).

SOLUCION. Después de la primera prueba en la urna quedaron solamente 5 bolillas, de ellas 3 blancas. La probabilidad condicional buscada es

$$P_A(B) = \frac{3}{5}.$$

Nota. De la definición de sucesos independientes se deduce que la aparición de uno de ellos no altera la probabilidad de que aparezca el otro. Por eso, para los sucesos independientes se cumplen las igualdades:

$$P_A(B) = P(B) \quad \text{y} \quad P_B(A) = P(A).$$

Por consiguiente, las probabilidades condicionales de sucesos independientes son iguales a las probabilidades absolutas.

§ 5. Teorema del producto de probabilidades de sucesos dependientes

Supongamos que los sucesos A y B son dependientes; además, las probabilidades $P(A)$ y $P_A(B)$ son conocidas. ¿Cómo hallar la probabilidad de simultaneidad de estos sucesos, es decir, la probabilidad de que aparezcan tanto el suceso A como el suceso B? La respuesta la da el teorema del producto.

Teorema. *La probabilidad de aparición simultánea de dos sucesos dependientes es igual al producto de la probabilidad de uno de ellos por la probabilidad condicional del otro calculada suponiendo que el primer suceso ha ocurrido ya*

$$P(AB) = P(A) \cdot P_A(B).$$

DEMOSTRACION. Introduzcamos las designaciones:

n, número de resultados elementales posibles de la prueba, en los cuales el suceso A ocurre o no;

n_1, número de resultados que son favorables al suceso A $(n_1 \leqslant n)$;

m, número de resultados elementales de la prueba, en los cuales ocurre el suceso B, suponiendo que el suceso A ya ocurrió, es decir, estos resultados favorecen a la producción del suceso AB $(m \leqslant n_1)$.

La probabilidad de que aparezcan simultáneamente los sucesos A y B es

$$P(AB)=\frac{m}{n}=\frac{n_1}{n}\cdot\frac{m}{n_1}.$$

Teniendo en cuenta que $\frac{n_1}{n}=P(A)$ y $\frac{m}{n_1}=P_A(B)$, finalmente obtenemos

$$P(AB)=P(A)\cdot P_A(B). \qquad (*)$$

Nota 1. Aplicando la fórmula (*) al suceso BA, tendremos: $P(BA)=P(B)\cdot P_B(A)$, o bien (puesto que el suceso BA no se diferencia del suceso AB)

$$P(AB)=P(B)\cdot P_B(A) \qquad (**)$$

Comparando las fórmulas (*) y (**), deducimos que se satisface la igualdad

$$P(A)\cdot P_A(B)=P(B)\cdot P_B(A). \qquad (***)$$

Corolario. *La probabilidad de que aparezcan simultáneamente varios sucesos dependientes es igual al producto de la probabilidad de uno de ellos por las probabilidades condicionales de todos los demás y además, la probabilidad de cada suceso sucesivo se calcula suponiendo que todos los sucesos anteriores han ocurrido ya:*

$$P(A_1A_2A_3\ldots A_n)=P(A_1)\cdot P_{A_1}(A_2)\cdot P_{A_1A_2}(A_3)\ldots$$
$$\ldots P_{A_1A_2\ldots A_{n-1}}(A_n),$$

donde $P_{A_1A_2\ldots A_{n-1}}(A_n)$ es la probabilidad del suceso A_n, calculada suponiendo que los sucesos $A_1, A_2, \ldots, A_{n-1}$ han ocurrido.

En particular, para tres sucesos dependientes tendremos:

$$P(ABC)=P(A)\cdot P_A(B)\cdot P_{AB}(C).$$

Cabe hacer notar que el orden, en el que se disponen los sucesos puede ser elegido libremente, es decir, es indiferente cuál suceso se considera primero, segundo, etc.

Para n arbitrario, la demostración se realiza por el método de inducción matemática.

Ejemplo 1. Un ajustador tiene 3 ejes cónicos y 7 elípticos. El ajustador toma al azar un eje y luego un segundo. Hallar la probabilidad de que el primer eje escogido es cónico y el segundo, elíptico.

SOLUCION. La probabilidad de que el primero de los ejes escogidos resulte cónico (suceso A) es

$$P(A)=\frac{3}{10}.$$

La probabilidad de que el segundo de los ejes sea elíptico (suceso B), calculada suponiendo que el primer eje es cónico, es decir, la probabilidad condicional es igual a

$$P_A(B)=\frac{7}{9}.$$

Por el teorema del producto de probabilidades de sucesos dependientes, la probabilidad buscada es igual a

$$P(AB)=P(A)\cdot P_A(B)=\frac{3}{10}\cdot\frac{7}{9}=\frac{7}{30}.$$

Observemos que, conservando la notación, hallamos fácilmente $P(B)=\frac{7}{10}$, $P_B(A)=\frac{3}{9}$, $P(B)\cdot P_B(A)=\frac{7}{30}$, lo que ilustra claramente que se cumple la igualdad (***).

Ejemplo 2. En una urna hay 5 bolillas blancas, 4 negras y 3 azules. En cada prueba se extrae al azar una bolilla, sin restituirla a la urna. Hallar la probabilidad de que en la primera prueba aparezca una bolilla blanca (suceso A), en la segunda, una negra (suceso B) y en la tercera, una azul (suceso C).

SOLUCION. La probabilidad de que aparezca una bolilla blanca en la primera prueba es

$$P(A)=\frac{5}{12}.$$

La probabilidad de que aparezca una bolilla negra en la segunda prueba, calculada suponiendo que en la primera prueba apareció una bolilla blanca, es decir, la probabilidad condicional es

$$P_A(B)=\frac{4}{11}.$$

La probabilidad de que aparezca una bolilla azul en la tercera prueba, calculada suponiendo que en la primera prueba apareció una bolilla blanca, y en la segunda, una negra, es

$$P_{AB}(C)=\frac{3}{10}.$$

La probabilidad buscada es

$$P(ABC) = P(A) \cdot P_A(B) \cdot P_{AB}(C) =$$
$$= \frac{5}{12} \cdot \frac{4}{11} \cdot \frac{3}{10} = \frac{1}{22} .$$

Nota 2. Expresemos la probabilidad condicional de la correlación (*) considerando $P(A) \neq 0$:

$$P_A(B) = \frac{P(AB)}{P(A)} .$$

Esta igualdad se puede tomar como definición de la probabilidad condicional.

Problemas

1. La probabilidad de que un tirador en un disparo haga blanco, es igual a $p = 0,9$. El tirador hizo 3 disparos. Hallar la probabilidad de que los tres disparos dieron al blanco.

Respuesta 0,729.

2. Se arrojan una moneda y un dado. Hallar la probabilidad de simultaneidad de los sucesos: «apareció la cara», «apareció el 6».

Respuesta $\frac{1}{12}$.

3. En dos cajas se encuentran piezas: en la primera, 10 (de ellas 3 standards), en la segunda, 15 (de ellas 6 standards). De cada caja se extrae al azar una pieza. Hallar la probabilidad de que ambas piezas resulten standards.

Respuesta 0,12.

4. En un estudio de televisión hay 3 cámaras de televisión. Para cada cámara la probabilidad de que ella esté conectada en el instante dado, es igual a $p = 0,6$. Hallar la probabilidad de que en el instante dado esté conectada por lo menos una cámara (suceso A).
Respuesta 0,936.

5. ¿A qué es igual la probabilidad de que al tirar tres dados aparezcan 6 puntos por lo menos en uno de los dados (suceso A)?

Respuesta $\frac{91}{216}$.

6. Una empresa produce 95% de artículos standards, además, de ellos 86%, de primera calidad. Hallar la probabilidad de que un artículo escogido al azar fabricado en esta empresa resulte de primera clase.

Respuesta 0,817.

7. Se arroja una moneda tantas veces hasta que no aparezcan 2 veces seguidas el mismo lado. Hallar la probabilidad de los sucesos siguientes: a) la prueba termina hasta la sexta tentativa; b) se necesita un número par de tentativas.

Respuesta a) $\frac{15}{16}$; b) $\frac{2}{3}$.

8. De las cifras 1, 2, 3, 4, 5 al principio elegimos una, y después de las cuatro restantes, una segunda cifra. Se supone que los 20 resultados posibles son equiprobables. Hallar la probabilidad de que será elegida una cifra impar: a) en la primera tentativa; b) en la segunda tentativa; c) en ambas tentativas.

Respuesta a) $\frac{3}{5}$; b) $\frac{3}{5}$; c) $\frac{3}{10}$.

9. La probabilidad de que en un disparo el tirador haga impacto en el diez, es igual a 0,6; ¿Cuántos disparos debe hacer el tirador, para que con una probabilidad no menor de 0,8 haga impacto en el diez por lo menos una vez?

Respuesta $n \geqslant 2$

10. Tres lámparas eléctricas están conectadas en serie al circuito. La probabilidad de que una (cualquiera) lámpara se encienda, cuando la tensión de la red supera a la nominal, es igual a 0,6. Hallar la probabilidad de que a una tensión elevada no habrá corriente en el circuito.

Respuesta 0,936.

11. La probabilidad de que el suceso A ocurra por lo menos una vez en dos pruebas independientes, es igual a 0,75. Hallar la probabilidad de que ocurra el suceso en una prueba (se supone que la probabilidad de que ocurra el suceso en ambas tentativas es la misma).

Respuesta 0,5

12. Tres equipos A_1, A_2, A_3 de la sociedad deportiva A compiten respectivamente con tres equipos de la sociedad B. Las probabilidades de que los equipos de la sociedad A ganen los juegos a los equipos de la sociedad B son: en el encuentro de A_1 con B_1, 0,8; de A_2 con B_2, 0,4; de A_3 con B_3, 0,4. Para el triunfo hay que ganar no menos de dos juegos de tres (los empates no se cuentan). ¿Cuál es la sociedad que tiene probabilidad de triunfar?

Respuesta. La sociedad A $\left(P_A = 0{,}544 > \frac{1}{2}\right)$.

13. La probabilidad de que el primer tirador haga blanco en un disparo es igual a 0,8 y el segundo tirador, 0,6. Hallar la probabilidad de que el blanco será hecho en un sólo disparo.

Respuesta 0,44.

14. De la sucesión de números 1, 2, . . ., n se eligen sucesivamente dos números al azar. Hallar la probabilidad de que uno de ellos es menor que un número entero positivo k y el otro mayor que k, donde $1 < k < n$.

Respuesta $\frac{2(k-1)(n-k)}{n(n-1)}$.

Advertencia. Admítase que: a) el primer número $< k$ y el segundo $> k$; b) el primer número $> k$, el segundo $< k$.

15. La sección de control técnico verifica el standard de los artículos. La probabilidad de que un artículo (producto) no es standard, es igual a 0,1. Hallar la probabilidad de que: a) de tres artículos verificados sólo uno resulta no standard; b) no standard resulta solamente el cuarto en orden de artículo controlado.

Respuesta a) 0,243; b) 0,0729.

Capítulo cuarto

COROLARIOS DE LOS TEOREMAS DE LA ADICION Y DEL PRODUCTO

§ 1. Teorema de la adición de probabilidades de sucesos simultáneos

Ya hemos visto el teorema de la adición para sucesos que se excluyen mutuamente. Aquí expondremos el teorema de la adición para sucesos simultáneos.

Dos sucesos se llaman *simultáneos*, si la aparición de uno de ellos no excluye la aparición del otro en una misma prueba.

Ejemplo. A es la aparición de cuatro puntos al tirar un dado; B es la aparición de un número par de puntos. Los sucesos A y B son simultáneos.

Supongamos que los sucesos A y B son simultáneos, además se conocen las probabilidades de estos sucesos y la probabilidad de su aparición simultánea. ¿Cómo hallar la probabilidad del suceso $A + B$, consistente en que ocurra por lo menos uno de los sucesos A y B? La respuesta la da el teorema de la adición de las probabilidades de sucesos simultáneos.

Teorema. *La probabilidad de que ocurra por lo menos uno de los dos sucesos simultáneos es igual a la suma de las probabi-*

lidades de estos sucesos sin la probabilidad de que ocurran simultáneamente

$$P(A+B)=P(A)+P(B)-P(AB).$$

DEMOSTRACION. Puesto que los sucesos A y B se suponen simultáneos, el suceso $A+B$ se produce, si ocurre uno de los tres sucesos que se excluyen mutuamente siguientes: $A\bar{B}$, $\bar{A}B$ o bien AB. Por el teorema de la adición de probabilidades de sucesos mutuamente excluyentes

$$P(A+B)=P(A\bar{B})+P(\bar{A}B)+P(AB) \qquad (*)$$

El suceso A ocurrirá, si se produce uno de los dos sucesos mutuamente excluyentes: $A\bar{B}$ o AB. Por el teorema de la adición de las probabilidades de sucesos que se excluyen mutuamente tendremos:

$$P(A)=P(A\bar{B})+P(AB).$$

De donde

$$P(A\bar{B})=P(A)-P(AB). \qquad (**)$$

Análogamente tendremos:

$$P(B)=P(\bar{A}B)+P(AB).$$

De aquí

$$P(\bar{A}B)=P(B)-P(AB). \qquad (***)$$

Sustituyendo (**) y (***) en (*), finalmente obtenemos:

$$P(A+B)=P(A)+P(B)-P(AB). \qquad (****)$$

Nota 1. Al utilizar las fórmulas obtenidas hay que tener en cuenta que los sucesos A y B pueden ser tanto independientes, como dependientes.

Para los sucesos independientes

$$P(A+B)=P(A)+P(B)-P(A)\cdot P(B);$$

para los sucesos dependientes

$$P(A+B)=P(A)+P(B)-P(A)\cdot P_A(B).$$

Nota 2. Si los sucesos A y B se excluyen mutuamente, su simultaneidad es un suceso incierto y, por lo tanto, $P(AB)=0$. La fórmula (****) para los sucesos mutuamente excluyentes toma la forma $P(A+B)=P(A)+P(B)$.

Nuevamente hemos obtenido el teorema de la adición para suce-

sos recíprocamente excluyentes. Así que la fórmula (****) se cumple tanto para los sucesos simultáneos, como para los alternativos o mutuamente excluyentes.

Ejemplo. Las probabilidades de hacer blanco al disparar el primer y segundo cañones son respectivamente iguales a: $p_1 = 0{,}7$; $p_2 = 0{,}8$. Hallar la probabilidad de impacto en un disparo (de ambos cañones) por *lo* menos de uno de los cañones.

SOLUCIÓN. La probabilidad de hacer blanco con cada uno de los cañones no depende del resultado del tiro de la otra arma, por eso los sucesos A (impacto de la primera arma) y B (impacto de la segunda arma) son independientes.

La probabilidad del suceso AB (ambas armas hicieron impacto) es

$$P(AB) = P(A) \cdot P(B) = 0{,}7 \cdot 0{,}8 = 0{,}56.$$

La probabilidad buscada es

$$P(A + B) = P(A) + P(B) - P(AB) =$$
$$= 0{,}7 + 0{,}8 - 0{,}56 = 0{,}94.$$

Nota. Puesto que en este ejemplo los sucesos A y B son independientes, se podría utilizar la fórmula $P = 1 - q_1q_2$ (cap. III, § 3). En realidad, las probabilidades de los sucesos que son opuestos a los sucesos A y B, es decir, las probabilidades de fallo son:

$$q_1 = 1 - p_1 = 1 - 0{,}7 = 0{,}3;$$
$$q_2 = 1 - p_2 = 1 - 0{,}8 = 0{,}2.$$

La probabilidad buscada de que en una descarga por lo menos una arma haga impacto, es igual a

$$P = 1 - q_1q_2 = 1 - 0{,}3 \cdot 0{,}2 = 0{,}94.$$

Como era de esperar, hemos obtenido el mismo resultado.

§ 2. Fórmula de la probabilidad completa

Supongamos que el suceso A puede ocurrir en condiciones de aparición de uno de los sucesos mutuamente excluyentes $B_1, B_2, \ldots, B_n$, que forman un grupo completo. Demos por conocidas las probabilidades de estos sucesos y las probabilidades condicionales $P_{B_1}(A)$, $P_{B_2}(A)$, $\ldots$, $P_{B_n}(A)$ del suceso A. ¿Cómo hallar la probabilidad del suceso A? La respuesta a esta pregunta la da el teorema siguiente.

Teorema. *La probabilidad del suceso A que puede ocurrir sólo a condición de que aparezca uno de los sucesos mutuamente excluyentes $B_1, B_2, \ldots, B_n$ que forman un grupo completo, es igual a la suma de los productos de las probabilidades de cada uno de estos sucesos por la correspondiente probabilidad condicional del suceso A:*

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + \ldots + P(B_n) \cdot P_{B_n}(A).$$

Esta fórmula, se llama «fórmula de la probabilidad completa».

DEMOSTRACION. Por hipótesis el suceso A puede ocurrir, si ocurre uno de los sucesos mutuamente excluyentes B_1, $B_2, \ldots, B_n$. En otras palabras, la aparición del suceso A significa la producción de uno, indiferentemente, de los sucesos mutuamente excluyentes $B_1A, B_2A, \ldots, B_nA$. Utilizando el teorema de la adición para el cálculo de la probabilidad del suceso A, obtenemos

$$P(A) = P(B_1A) + P(B_2A) + \ldots + P(B_nA). \quad (*)$$

Queda por calcular cada uno de los sumandos. Por el teorema del producto de las probabilidades de sucesos dependientes tendremos

$$P(B_1A) = P(B_1) \cdot P_{B_1}(A); \; P(B_2A) = P(B_2) \cdot P_{B_2}(A); \; \ldots; \; \ldots; \; P(B_nA) = P(B_n) \cdot P_{B_n}(A).$$

Sustituyendo los segundos miembros de estas igualdades en la correlación (*), obtenemos la fórmula de la probabilidad completa

$$P(A) = P(B_1) \cdot P_{B_1}(A) + P(B_2) \cdot P_{B_2}(A) + \ldots + P(B_n) \cdot P_{B_n}(A).$$

Ejemplo 1. Se tienen dos juegos de piezas. La probabilidad de que una pieza del primer juego sea standard, es igual a 0,8, y del segundo, 0,9. Hallar la probabilidad de que la pieza tomada al azar (de un juego escogido al acaso) sea standard.

SOLUCION. Designemos por A el suceso, la pieza extraída es standard.

La pieza puede ser extraída del primer juego (suceso B_1), o bien del segundo juego (suceso B_2).

La probabilidad de que la pieza será escogida del primer juego es

$$P(B_1)=\frac{1}{2}.$$

La probabilidad de que la pieza será tomada del segundo juego es

$$P(B_2)=\frac{1}{2}.$$

La probabilidad condicional de que del primer juego será extraída una pieza standard es

$$P_{B_1}(A)=0{,}8.$$

La probabilidad condicional de que del segundo juego será extraída una pieza standard es

$$P_{B_2}(A)=0{,}9.$$

La probabilidad buscada de que la pieza sacada al azar sea standard, por la fórmula de la probabilidad completa es igual a

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)=$$
$$=0{,}5\cdot 0{,}8+0{,}5\cdot 0{,}9=0{,}85.$$

Ejemplo 2. En la primera caja hay 20 lámparas de radio, de las cuales 18 son standard; en la segunda caja hay 10 válvulas, de las cuales 9 son standard. De la segunda caja se ha tomado una válvula al azar y se ha colocado en la primera. Hallar la probabilidad de que la válvula extraída al acaso de la primera caja sea standard.

SOLUCION. Designamos por A el suceso, de la primera caja se ha extraído una válvula standard

De la segunda caja pudo haberse sacado una válvula standard (suceso B_1) o no standard (suceso B_2).

La probabilidad de que de la segunda caja se ha extraído una válvula standard es

$$P(B_1)=\frac{9}{10}.$$

La probabilidad de que de la segunda caja se ha escogido una válvula no standard es

$$P(B_2)=\frac{1}{10}.$$

La probabilidad condicional de que de la primera caja se ha extraído una válvula standard, a condición de que de la segunda caja se traspasó a la primera una válvula standard, es

$$P_{B_1}(A)=\frac{19}{21}.$$

La probabilidad condicional de que de la primera caja se extrajo una válvula standard, a condición de que de la segunda caja se colocó en la primera una válvula no standard, es

$$P_{B_2}(A)=\frac{18}{21}.$$

La probabilidad buscada de que, de la primera caja será sacada una válvula standard, por la fórmula de probabilidad completa, es igual a

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)=$$
$$=\frac{9}{10}\cdot\frac{19}{21}+\frac{1}{10}\cdot\frac{18}{21}=0{,}9.$$

§ 3. Probabilidad de las hipótesis. Fórmula de Bayes

Supongamos que el suceso A puede ocurrir a condición de que aparezca uno de los sucesos mutuamente excluyentes B_1, B_2, . . ., B_n, que forman un grupo completo. Puesto que de antemano no se sabe cuál de estos sucesos ocurrirá, ellos se llaman *hipótesis*. La probabilidad de que aparezca el suceso A se determina por la fórmula de probabilidad completa (§ 2):

$$P(A)=P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+\ldots+$$
$$+P(B_n)\cdot P_{B_n}(A). \qquad (*)$$

Admitamos que se realizó un experimento, debido al cual ocurrió el suceso A. Nos planteamos determinar cómo han variado (a consecuencia de que el suceso A ya ocurrió) las probabilidades de las hipótesis. En otras palabras, vamos a buscar las probabilidades condicionales

$$P_A(B_1),\ P_A(B_2),\ \ldots,\ P_A(B_n).$$

Primero hallamos la probabilidad condicional $P_A(B_1)$. Por el teorema del producto tendremos

$$P(AB_1)=P(A)\cdot P_A(B_1)=P(B_1)\cdot P_{B_1}(A).$$

De donde

$$P_A(B_1)=\frac{P(B_1)\cdot P_{B_1}(A)}{P(A)}.$$

Sustituyendo aquí $P(A)$ por la fórmula (*), obtenemos

$$P_A(B_1)=\frac{P(B_1)\cdot P_{B_1}(A)}{P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+\ldots+P(B_n)\cdot P_{B_n}(A)}.$$

Análogamente se deducen las fórmulas que determinan las probabilidades condicionales de las demás hipótesis, es decir, la probabilidad condicional de cualquier hipótesis B_i ($i=1, 2, \ldots, n$) puede ser calculada por la fórmula

$$P_A(B_i)=\frac{P(B_i)\cdot P_{B_i}(A)}{P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)+\ldots+P(B_n)\cdot P_{B_n}(A)}.$$

Las fórmulas obtenidas se llaman *fórmulas de Bayes* (nombre del matemático inglés que las dedujo, siendo publicadas en 1764). *Las fórmulas de Bayes permiten volver a estimar las probabilidades de las hipótesis después de conocer el resultado de la experimentación, debido a la cual ocurrió el suceso* A.

Ejemplo. Las piezas producidas por una sección de la fábrica caen para su verificación de standard a uno de dos revisores. La verificación de que una pieza llega al primer revisor, es igual a 0,6, y al segundo, 0,4. La probabilidad de que la pieza acabada será reconocida como standard por el primer revisor es igual a 0,94 y por el segundo, 0,98. La pieza acabada ha sido considerada standard. Hallar la probabilidad de que esta pieza fue controlada por el primer revisor.

SOLUCION. Designemos por A el suceso constituido por el reconocimiento de la pieza acabada como standard. Se pueden hacer dos hipótesis:

1) la pieza fue controlada por el primer revisor (hipótesis B_1);

2) la pieza fue controlada por el segundo revisor (hipótesis B_2).

La probabilidad buscada de que la pieza fue controlada por el primer revisor, la hallamos por la fórmula de Bayes

$$P_A(B_1)=\frac{P(B_1)\cdot P_{B_1}(A)}{P(B_1)\cdot P_{B_1}(A)+P(B_2)\cdot P_{B_2}(A)}.$$

Según datos del problema tendremos:

$P(B_1) = 0{,}6$ (la probabilidad de que la pieza llega al primer revisor);

$P(B_2) = 0{,}4$ (la probabilidad de que la pieza llega al segundo revisor);

$P_{B_1}(A) = 0{,}94$ (la probabilidad de que la pieza acabada será reconocida como standard por el primer revisor):

$P_{B_2}(A) = 0{,}98$ (la probabilidad de que la pieza acabada será considerada como standard por el segundo revisor).

La probabilidad buscada es

$$P_A(B_1) = \frac{0{,}6 \cdot 0{,}4}{0{,}6 \cdot 0{,}94 + 0{,}4 \cdot 0{,}98} \approx 0{,}59.$$

Como se aprecia, hasta la prueba la probabilidad de la hipótesis B_1 ha sido igual a 0,6 y después de conocerse el resultado de la prueba, la probabilidad de esta hipótesis (más exactamente, la probabilidad condicional) cambió y se hizo igual a 0,59. De este modo, el empleo de la fórmula de Bayes permitió volver a estimar la probabilidad de la hipótesis considerada.

Problemas

1. Dos tiradores hicieron un disparo cada uno. La probabilidad de que el primer tirador haga blanco es igual a 0.7 y la del segundo, 0.6. Hallar la probabilidad de que por lo menos uno de los tiradores haya dado en el blanco.

Respuesta 0.88.

2. Un montador tiene 16 piezas producidas por la fábrica № 1 y 4 piezas de la fábrica № 2. Se han tomado 2 piezas al azar. Hallar la probabilidad de que por lo menos una de ellas resulte producida por la fábrica № 1.

Respuesta $\frac{92}{95}$.

3. En un grupo de deportistas 20 son esquiadores, 6 ciclistas y 4 corredores. La probabilidad de cumplir la prueba de cualificación es la siguiente: para el esquiador 0,9; para el ciclista 0,8 y para el corredor 0,75. Hallar la probabilidad de que el deportista escogido al acaso cumpla la prueba.

Respuesta 0,86.

4. Un montador recibió 3 cajas de piezas producidas por la fábrica № 1 y 2 cajas de piezas producidas por la fábrica № 2. La probabilidad de que la pieza de la fábrica № 1 sea standard es igual a 0,8 y la de la fábrica № 2, 0,9. El montador extrajo al acaso una pieza de una caja tomada al azar. Hallar la probabilidad de que se haya extraído una pieza standard.

Respuesta 0,84.

5. En la primera caja hay 20 piezas, de las cuales 15 son standard; en la segunda hay 30 piezas, de ellas 24 son standard; en la tercera, 10 piezas, de las cuales 6 son standard. Hallar la probabilidad de que la pieza escogida al acaso de una caja tomada al azar sea standard.

Respuesta $\frac{43}{60}$.

6. En un taller de televisión hay 4 cinescopios. Las probabilidades de que el cinescopio observe el plazo de garantía de funcionamiento, respectivamente, son iguales a 0,8; 0,85; 0,9; 0,95. Hallar la probabilidad de que el cinescopio tomado al azar observe el plazo de garantía de servicio.

Respuesta 0,875.

7. En dos cajas hay válvulas de radio. La primera caja contiene 12 válvulas, de ellas una no standard; en la segunda 10 válvulas, de ellas 1, no standard. De la primera caja se ha extraído al azar una válvula y transpasado a la segunda caja. Hallar la probabilidad de que la válvula tomada al azar de la segunda caja sea no standard.

Respuesta $\frac{13}{132}$.

8. De un juego completo de 28 fichas de dominó se ha extraído una al azar. Hallar la probabilidad de que la segunda ficha tomada al azar se pueda juntar a la primera.

Respuesta $\frac{7}{18}$.

9. Un estudiante no sabe todas las papeletas de examen. ¿En qué caso la probabilidad de escoger una papeleta desconocida será la menor cuando toma la primera o la última papeleta?

Respuesta. La probabilidad es idéntica en ambos casos.

10. En una caja que contiene 3 piezas idénticas, se ha tirado una pieza standard, y después al azar se ha extraído una pieza. Hallar la probabilidad de que se haya sacado la pieza standard, si son equiprobables todas hipótesis posibles sobre el número de piezas standard que inicialmente se encontraban en la caja.

Respuesta 0,625.

11. Cuando el automático se desvía del régimen de trabajo normal funciona el avisador *C*-1 con la probabilidad de 0,8, y el avisador *C*-11 funciona con la probabilidad de 1. La probabilidad de que el

automático está equipado del avisador C-1 o C-11 es igual a 0,6 y 0,4 respectivamente. Se ha recibido la señal de desperfecto del automático. ¿Qué es lo más probable: el automático está equipado del avisador C-1 o C-11?

Respuesta. La probabilidad de que el automático está equipado del avisador C-1 es igual a $\frac{6}{11}$, y C-11, $\frac{5}{11}$.

12. Para participar en las pruebas deportivas eliminatorias estudiantiles se han seleccionado del primer grupo del curso 4 estudiantes, del segundo grupo, 6 estudiantes; del tercer grupo, 5 estudiantes. Las probabilidades de que un estudiante del primer, segundo y tercer grupos participe en el seleccionado del instituto, son respectivamente iguales a 0,9; 0,7 y 0,8. El estudiante elegido al azar al final de la competición entró en el seleccionado. ¿A cuál de los grupos perteneció con más probabilidad este estudiante?

Respuesta. Las probabilidades de que se haya elegido un estudiante del primer, segundo, tercer grupos, son respectivamente iguales a:

$$\frac{18}{59}, \quad \frac{21}{59}, \quad \frac{20}{59}.$$

13. La probabilidad de que los artículos de cierta producción satisfagan al standard es igual a 0,96. Se propone el sistema simplificado de verificación del standard que da un resultado positivo con la probabilidad de 0,98 para los artículos que satisfacen el standard, y para los artículos que no satisfacen el standard, con probabilidad de 0,05. Hallar la probabilidad de que el artículo, reconocido como standard durante la verificación, satisface en realidad el standard.

Respuesta 0,998.

Capítulo quinto

REPETICION DE LOS EXPERIMENTOS

§ 1. Fórmula de Bernoulli

Si se realizan varios experimentos, además la probabilidad del suceso A en cada prueba no depende de los resultados de otras pruebas, tales experimentos se llaman *independientes con respecto al suceso A*.

En distintas pruebas independientes el suceso A puede tener diferentes probabilidades, o bien la misma probabilidad. En adelante consideraremos sólo estos experimentos independientes, en los cuales el suceso A tiene la misma probabilidad.

Más adelante utilizamos el concepto de suceso complejo o compuesto, sobreentendiendo por ello la simultaneidad de varios sucesos individuales, llamados simples.

Supongamos que se realizan n pruebas independientes, en cada una de las cuales el suceso A puede ocurrir o no ocurrir. Vamos a considerar que la probabilidad del suceso A en cada prueba es la misma y, precisamente, igual a p. Por lo tanto, la probabilidad de que el suceso A no ocurra en cada prueba también es constante e igual a $q = 1 - p$.

Tratemos de calcular la probabilidad de que en n pruebas el suceso A ocurra exactamente k veces y, por lo tanto, no ocurre $n - k$ veces.

Es importante señalar que no se exige que el suceso A se repita exactamente k veces en una sucesión determinada. Por ejemplo, si se trata de la aparición del suceso A tres veces en cuatro pruebas. son posibles los siguientes sucesos compuestos:

$$AAA\bar{A}, \quad AA\bar{A}A, \quad A\bar{A}AA \quad \text{y} \quad \bar{A}AAA.$$

La notación $AAA\bar{A}$ significa que en la primera, segunda y tercera pruebas o tentativas el suceso A ocurrió, mientras que en la cuarta tentativa no ocurrió, es decir, ocurrió un suceso opuesto $\bar{A}$. Las otras notaciones también tienen el significado correspondiente.

A la probabilidad buscada la designamos por $P_n(k)$. Por ejemplo, el símbolo $P_5(3)$ denota la probabilidad de que en cinco tentativas el suceso ocurra exactamente 3 veces y, por lo tanto, no ocurra 2 veces.

El problema planteado lo resuelve la fórmula de Bernoulli.

Deducción de la fórmula de Bernoulli. La probabilidad de un suceso compuesto consistente en que, en n tentativas el suceso A ocurra k veces y no ocurra $n - k$ veces, por el teorema del producto de probabilidades de sucesos independientes, es igual a

$$p^k q^{n-k}.$$

Estos sucesos compuestos pueden ser tantos, como las combinaciones posibles de n elementos de k elementos, es decir, C_n^k. Dado que estos sucesos compuestos se excluyen mutuamente, por el teorema de la adición de probabilidades de sucesos mutuamente excluyentes, la probabilidad buscada es igual a la suma de las probabilidades de todos los sucesos

compuestos posibles. Puesto que las probabilidades de todos estos sucesos compuestos son idénticas, la probabilidad buscada (de que el suceso A ocurra k veces en n tentativas) es igual a la probabilidad de un suceso compuesto, multiplicado por el número de ellos:

$$P_n(k) = C_n^k p^k q^{n-k},$$

o bien

$$P_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k}.$$

La ecuación obtenida se llama fórmula de Bernoulli.

Ejemplo. La probabilidad de que el consumo de energía eléctrica durante unos días no sea mayor que la norma establecida, es igual a $p = 0{,}75$. Hallar la probabilidad de que en los próximos 6 días el consumo de energía eléctrica durante 4 días no supere la norma.

SOLUCION. La probabilidad de consumo normal de energía eléctrica durante cada uno de los 6 días es constante e igual a $p = 0{,}75$. Por lo tanto, la probabilidad de un sobreconsumo de energía en cada día también es constante e igual a $q = 1 - p = 1 - 0{,}75 = 0{,}25$.

Por la fórmula de Bernoulli la probabilidad buscada es

$$P_6(4) = C_6^4 p^4 q^2 = C_6^2 p^4 q^2 = \frac{6 \cdot 5}{1 \cdot 2} (0{,}75)^4 \cdot (0{,}25)^2 = 0{,}30.$$

§ 2. Teorema local de Laplace

En el párrafo anterior deducimos la fórmula de Bernoulli que permite calcular la probabilidad de que el suceso ocurrirá en n tentativas exactamente k veces. En la deducción hemos supuesto que la probabilidad de que ocurra un suceso en cada tentativa es constante.

Se aprecia fácilmente que es bastante difícil utilizar la fórmula de Bernoulli para grandes valores de n, ya que hay que operar con números colosales. Por ejemplo, si $n = 50$, $k = 30$, $p = 0{,}1$, para hallar la probabilidad $P_{50}(30)$ hay que calcular la expresión $P_{50}(30) = \frac{50!}{30!\,20!} (0{,}1)^{30} \cdot (0{,}9)^{20}$, donde $50! = 30\,414\,093 \cdot 10^{57}$, $30! = 26\,525\,286 \cdot 10^{25}$, $20! = 24\,329\,020 \cdot 10^{11}$. Claro está que se puede simplificar algo el cálculo, utilizando tablas especiales de logaritmos de facto-

riales. Sin embargo, también este camino sigue siendo voluminoso y, por lo mismo, presenta un importante inconveniente: las tablas contienen valores aproximados de los logaritmos, por lo cual en los cálculos se acumulan errores; en suma, el resultado final puede diferenciarse bastante del verdadero.

Naturalmente surge la pregunta: ¿acaso no se podría calcular la probabilidad que nos interesa sin recurrir a la fórmula de Bernoulli? Resulta que sí, se puede. El teorema local de Laplace da precisamente la fórmula asintótica* que permite hallar aproximadamente la probabilidad que ocurra un suceso exactamente k veces en n tentativas, si el número de tentativas o pruebas es bastante grande.

Cabe hacer notar que, para el caso particular, precisamente para $p=\frac{1}{2}$, la fórmula asintótica fue descubierta en 1730 por Moivre; en el año 1783 Laplace generalizó la fórmula de Moivre para p arbitrario, distinto de 0 y de 1. Por eso, el teorema, al que nos referimos, suele llamarse teorema de Moivre — Laplace.

La demostración del teorema local de Laplace es bastante compleja, por lo cual sólo formularemos el teorema y daremos ejemplos que ilustran su aplicación.

Teorema local de Laplace. *Si p es la probabilidad de que ocurra un suceso A en cada tentativa es constante y diferente de cero y de la unidad, la probabilidad $P_n(k)$ de que el suceso A ocurra en n tentativas exactamente k veces, aproximadamente es igual (tanto más exacta, cuanto mayor es n) al valor de la función*

$$y=\frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\frac{1}{\sqrt{npq}}\cdot\varphi(x),$$

para $x=\frac{k-np}{\sqrt{npq}}$.

Existen tablas en los que se dan los valores de función $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, correspondientes a los valores positivos del argumento x (suplemento 1). Para los valores negativos del argumento se utilizan las mismas tablas, ya que la función $\varphi(x)$ es par, es decir, $\varphi(-x)=\varphi(x)$.

* La función $\varphi(x)$ se llama función aproximada asintóticamente $f(x)$, si $\lim_{x\to\infty}\frac{f(x)}{\varphi(x)}=1$.

De este modo, la probabilidad de que un suceso A ocurra en n tentativas independientes exactamente k veces, aproximadamente es igual a

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x),$$

donde $x = \dfrac{k-np}{\sqrt{npq}}$.

Ejemplo 1. Hallar la probabilidad de que el suceso A ocurrirá exactamente 80 veces en 400 tentativas, si la probabilidad de que ocurra este suceso en cada tentativa es igual a 0,2.

SOLUCION. Por los datos del problema $n = 400$; $k = 80$; $P = 0{,}2$; $q = 0{,}8$. Utilizamos la fórmula asintótica de Laplace:

$$P_{400}(80) \approx \frac{1}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} \cdot \varphi(x) = \frac{1}{8} \cdot \varphi(x).$$

Calculemos el valor de x, determinable por los datos del problema

$$x = \frac{k-np}{\sqrt{npq}} = \frac{80 - 400 \cdot 0{,}2}{8} = 0.$$

Por la tabla (suplemento 1) hallamos $\varphi(0) = 0{,}3989$.

La probabilidad buscada es

$$P_{400}(80) = \frac{1}{8} \cdot 0{,}3989 = 0{,}04986.$$

La fórmula de Bernoulli nos conduce aproximadamente al mismo resultado (debido a su voluminosidad, se han omitido los cálculos):

$$P_{400}(80) = 0{,}0498.$$

Ejemplo 2. La probabilidad de que un tirador falle al blanco en un solo disparo es $p = 0{,}75$. Hallar la probabilidad de que en 10 disparos el tirador logre hacer blanco 8 veces.

SOLUCION. Por los datos $n = 10$; $k = 8$; $p = 0{,}75$; $q = 0{,}25$. Utilizamos la fórmula asintótica de Laplace:

$$P_{10}(8) \simeq \frac{1}{\sqrt{10 \cdot 0{,}75 \cdot 0{,}25}} \cdot \varphi(x) = 0{,}7301 \cdot \varphi(x).$$

Calculemos el valor de x, determinable por los datos del problema:

$$x = \frac{k - np}{\sqrt{npq}} = \frac{8 - 10 \cdot 0,75}{\sqrt{10 \cdot 0,75 \cdot 0,25}} \simeq 0,36.$$

Por la tabla (suplemento 1) hallamos $\varphi(0,36) = 0,3739$

La probabilidad buscada es

$$P_{10}(8) = 0,7301 \cdot 0,3739 = 0,273.$$

La fórmula de Bernoulli conduce a otro resultado, precisamente a $P_{10}(8) = 0,282$. La diferencia tan grande de los resultados se debe a que, en este ejemplo n tiene un valor pequeño (la fórmula de Laplace da una aproximación bastante buena sólo para valores de n suficientemente grandes).

§ 3. Teorema integral de Laplace

Nuevamente supongamos que se realizan n pruebas, en cada una de las cuales la probabilidad de que ocurra el suceso A es constante e igual a p $(0 < p < 1)$. ¿Cómo calcular la probabilidad $P_n(k_1, k_2)$, de que en n tentativas el suceso A ocurrirá no menos de k_1 y no más de k_2 veces (para abreviar diremos «desde k_1 hasta k_2 veces»)? La respuesta a esta pregunta la da el teorema integral de Laplace que damos a continuación, omitiendo la demostración.

Teorema. *Si la probabilidad p de que ocurra un suceso A en cada tentativa es constante y distinta de cero y de la unidad, la probabilidad $P_n(k_1, k_2)$ de que el suceso A ocurra en n tentativas desde k_1 hasta k_2 veces, aproximadamente es igual a una integral definida*

$$P_n(k_1, k_2) \simeq \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-\frac{z^2}{2}} \, dz, \qquad (*)$$

donde

$$x' = \frac{k_1 - np}{\sqrt{npq}} \quad \text{y} \quad x'' = \frac{k_2 - np}{\sqrt{npq}}.$$

Al resolver los problemas que requieren la aplicación del teorema integral de Laplace, se utilizan tablas especiales, puesto que la integral indefinida $\int e^{-\frac{z^2}{2}} \, dz$ no se expresa por

funciones elementales. Al final del libro se da la tabla (suplemento 2) para la integral $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz$. En la tabla se dan los valores de la función $\Phi(x)$ para x positivos y para $x=0$; para $x<0$ se utiliza la misma tabla (la función $\Phi(x)$ es impar, es decir, $\Phi(-x)=-\Phi(x)$). En la tabla se dan los valores de la integral sólo hasta $x=5$, ya que para $x>5$ se puede tomar $\Phi(x)=0,5$. De ordinario $\Phi(x)$ se llama función de Laplace.

Para poder utilizar la tabla de la función de Laplace, transformamos la correlación (*) así:

$$P_n(k_1,\ k_2)\simeq\frac{1}{\sqrt{2\pi}}\int_{x'}^{0} e^{-\frac{z^2}{2}}dz+\frac{1}{\sqrt{2\pi}}\int_0^{x''} e^{-\frac{z^2}{2}}dz=$$

$$=\frac{1}{\sqrt{2\pi}}\int_0^{x''} e^{-\frac{z^2}{2}}dz-\frac{1}{\sqrt{2\pi}}\int_0^{x'} e^{-\frac{z^2}{2}}dz=\Phi(x'')-\Phi(x').$$

De este modo, la probabilidad de que el suceso A ocurre en n tentativas independientes desde k_1 hasta k_2 veces

$$P_n(k_1,\ k_2)\simeq\Phi(x'')-\Phi(x'),$$

donde $x'=\frac{k_1-np}{\sqrt{npq}}$ y $x''=\frac{k_2-np}{\sqrt{npq}}$.

Veamos ejemplos que ilustran la aplicación del teorema integral de Laplace.

Ejemplo. La probabilidad de que una pieza no haya pasado la comprobación de la SCT, es igual a $p=0,2$. Hallar la probabilidad de que entre 400 piezas escogidas fortuitamente resulten no controladas desde 70 hasta 100 piezas.

SOLUCION. Por los datos del problema $p=0,2$; $q=0,8$; $n=400$; $k_1=70$; $k_2=100$.

Utilizamos el teorema integral de Laplace

$$P_{400}(70,\ 100)\approx\Phi(x'')-\Phi(x').$$

Calculemos los límites inferior y superior de la integración:

$$x'=\frac{k_1-np}{\sqrt{npq}}=\frac{70-400\cdot0,2}{\sqrt{400\cdot0,2\cdot0,8}}=-1,25;$$

$$x''=\frac{k_2-np}{\sqrt{npq}}=\frac{100-400\cdot0,2}{\sqrt{400\cdot0,2\cdot0,8}}=2,5.$$

De este modo, tendremos

$$P_{400}(70,\ 100) = \Phi(2{,}5) - \Phi(-1{,}25) = \Phi(2{,}5) + \Phi(1{,}25).$$

Por la tabla (suplemento 2) hallamos

$$\Phi(2{,}5) = 0{,}4938; \qquad \Phi(1{,}25) = 0{,}3944.$$

La probabilidad buscada es

$$P_{400}(70,\ 100) = 0{,}4938 + 0{,}3944 = 0{,}8882.$$

Nota. Designemos por m el número de apariciones del suceso A para n tentativas independientes, en cada una de las cuales la probabilidad de que ocurra el suceso A es constante e igual a p. Si el número m varía desde k_1 hasta k_2, tendremos que la fracción $\frac{m-np}{\sqrt{npq}}$ variará desde $\frac{k_1-np}{\sqrt{npq}}=x'$ hasta $\frac{k_2-np}{\sqrt{npq}}=x''$. Por lo tanto, el teorema integral de Laplace también se puede escribir así

$$P\left(x' \leqslant \frac{m-np}{\sqrt{npq}} \leqslant x''\right) \simeq \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-\frac{z^2}{2}}\, dz.$$

Esta forma de notación la utilizamos más adelante.

§ 4. Probabilidad de desviación de la frecuencia relativa respecto de la probabilidad constante en experimentos independientes

Consideraremos nuevamente que se realizan n experimentos independientes, en cada uno de los cuales la probabilidad de que aparezca el suceso A es constante e igual a p $(0 < p < 1)$. Tratemos de hallar la probabilidad de que la desviación de la frecuencia relativa $\frac{m}{n}$ respecto de la probabilidad constante p, no es mayor, en valor absoluto, que un número dado $\varepsilon > 0$. En otras palabras, hallamos la probabilidad de que se cumple la desigualdad

$$\left|\frac{m}{n} - p\right| \leqslant \varepsilon. \qquad (*)$$

Esta probabilidad la designaremos así: $P\left(\left|\frac{m}{n}-p\right|\leqslant\varepsilon\right)$.
Sustituimos la desigualdad (*) por sus equivalentes:

$$-\varepsilon\leqslant\frac{m}{n}-p\leqslant\varepsilon, \quad \text{o bien} \quad -\varepsilon\leqslant\frac{m-np}{n}\leqslant\varepsilon.$$

Multiplicando estos desigualdades por el factor positivo $\sqrt{\frac{n}{pq}}$, obtenemos las desigualdades equivalentes a la inicial:

$$-\varepsilon\sqrt{\frac{n}{pq}}\leqslant\frac{m-np}{\sqrt{npq}}\leqslant\varepsilon\sqrt{\frac{n}{pq}}.$$

Aplicamos el teorema integral de Laplace a la fórmula dada en la nota de la pag. 64. Poniendo

$$x'=-\varepsilon\sqrt{\frac{n}{pq}} \quad \text{y} \quad x''=\varepsilon\sqrt{\frac{n}{pq}}, \text{ tendremos:}$$

$$P\left(-\varepsilon\sqrt{\frac{n}{np}}\leqslant\frac{m-np}{\sqrt{npq}}\leqslant\varepsilon\sqrt{\frac{n}{nq}}\right)\approx$$

$$\approx\frac{1}{\sqrt{2\pi}}\int\limits_{-\varepsilon\sqrt{\frac{n}{pq}}}^{\varepsilon\sqrt{\frac{n}{pq}}} e^{-\frac{z^2}{2}}\,dz=$$

$$=\frac{2}{\sqrt{2\pi}}\int\limits_{0}^{\varepsilon\sqrt{\frac{n}{pq}}} e^{-\frac{z^2}{2}}\,dz=2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Por último, sustituyendo la desigualdad por la desigualdad inicial equivalente a ella, cerrada entre paréntesis, finalmente obtenemos:

$$P\left(\left|\frac{m}{n}-p\right|\leqslant\varepsilon\right)\simeq 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

De este modo, la probabilidad de que se cumpla la desigualdad

$$\left|\frac{m}{n}-p\right|\leqslant\varepsilon$$

aproximadamente es igual al valor doble de la función de Laplace $2\Phi(x)$ para $x=\varepsilon\sqrt{\frac{n}{pq}}$.

Ejemplo 1. La probabilidad de que una pieza no es standard, es $p = 0{,}1$. Hallar la probabilidad de que entre 400 piezas escogidas fortuitamente la frecuencia relativa de aparición de piezas no standard se desvía de la probabilidad $p = 0{,}1$, en valor absoluto, no más de 0,03.

SOLUCION. Por los datos del problema $n = 400$; $p = 0{,}1$; $q = 0{,}9$; $\varepsilon = 0{,}03$.

Se requiere hallar la probabilidad $P\left(\left|\frac{m}{400} - 0{,}1\right| \leqslant 0{,}03\right)$.

Utilizando la fórmula $P\left(\left|\frac{m}{n} - p\right| \leqslant \varepsilon\right) \approx$ $\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$, tendremos: $P\left(\left|\frac{m}{400} - 0{,}1\right| \leqslant 0{,}03\right) \approx$ $\approx 2\Phi\left(0{,}03\sqrt{\frac{400}{0{,}1 \cdot 0{,}9}}\right) = 2\Phi(2)$.

Por la tabla (suplemento 2) hallamos $\Phi(2) = 0{,}4772$. Por lo tanto, $2\Phi(2) = 0{,}9544$.

De este modo, la probabilidad buscada es aproximadamente igual a 0,9544.

Este resultado nos dice que: si se toma un número suficientemente grande de pruebas de a 400 piezas cada una, aproximadamente en el 95,44% de estas pruebas la desviación de la frecuencia relativa respecto de la probabilidad constante $P = 0{,}1$ no supera en valor absoluto 0,03.

Ejemplo 2. La probabilidad de que la pieza no es standard, es $p = 0{,}1$. Hallar la cantidad de piezas que se debe escoger para que con una probabilidad igual a 0,9544 se pudiese confirmar que la frecuencia relativa de aparición de piezas no standards (entre las escogidas) se desvía de la probabilidad constante p no más de 0,03 en valor absoluto.

SOLUCION. Por los datos del problema $p = 0{,}1$; $q = 0{,}9$; $\varepsilon = 0{,}03$;

$$P\left(\left|\frac{m}{n} - 0{,}1\right| \leqslant 0{,}03\right) = 0{,}0544.$$

Hay que hallar n.

Utilizamos la fórmula $P\left(\left|\frac{m}{n} - p\right| \leqslant \varepsilon\right) \approx$ $\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$.

Según los datos,

$$2\Phi\left(0{,}03\sqrt{\frac{n}{0{,}1 \cdot 0{,}9}}\right) = 2\Phi(0{,}1\sqrt{n}) = 0{,}9544.$$

Por lo tanto, $\Phi(0{,}1\sqrt{n}) = 0{,}4772$.

Por la tabla (suplemento 2) hallamos

$$\Phi(2) = 0,4772.$$

Para hallar el número n obtenemos la ecuación

$$0,1\sqrt{n} = 2.$$

De donde el número de piezas buscado $n = 400$.

El resultado obtenido nos dice que: si se toma un número suficientemente grande de pruebas de a 400 piezas, en el 95,44% de estas pruebas la frecuencia relativa de aparición de piezas no standards se diferencia de la probabilidad constante $p = 0,1$ no más de 0,03 en valor absoluto, es decir, la frecuencia relativa estará comprendida entre los límites desde 0,07 ($0,1 - 0,03 = 0,07$) hasta 0,13 ($0,1 + 0,03 = 0,13$).

En otras palabras, el número de piezas no standard en el 95,44% de las pruebas estará comprendido entre 28 (7% de 400) y 52 (13% de 400).

Si se toma solamente una prueba de 400 piezas, con gran certeza se puede esperar que en esa prueba habrá no menos de 28 y no más de 52 piezas no standard. Es posible, aunque poco probable, que las piezas no standards resulten menos de 20 o bien más de 52.

Problemas

1. En un taller hay 6 motores. La probabilidad para cada motor de que en el instante dado esté conectado es igual a 0,8. Hallar la probabilidad de que en el instante dado: a) están conectados 4 motores; b) están conectados todos los motores; c) todos los motores están desconectados.

Respuesta a) $P_6(4) = 0,246$; b) $P_6(6) = 0,26$; c) $P_6(0) = 0,000064$.

2. Hallar la probabilidad de que el suceso A ocurrirá no menos de dos veces en cinco pruebas independientes, si en cada tentativa la probabilidad de que ocurra el suceso A es igual a 0,3.

Respuesta $P = 1 - [P_5(0) + P_5(1)] = 0,472.$

3. El suceso B ocurrirá, si el suceso A se produce no menos de dos veces. Hallar la probabilidad de que el suceso B ocurrirá, si se realizan 6 pruebas independientes, en cada una de las cuales la probabilidad de que ocurra el suceso A es igual a 0,4.

Respuesta $P = 1 - [P_6(0) + P_6(1)] = 0,767.$

4. Se han realizado 8 experimentos independientes, en cada uno de los cuales la probabilidad de ocurrir el suceso A es igual a 0,1. Hallar la probabilidad de que el suceso A ocurra por lo menos 2 veces.

Respuesta $P = 1 - [P_8(0) + P_8(1)] = 0{,}19$.

5. Se arroja una moneda 6 veces. Hallar la probabilidad de que caiga cara: a) menos de dos veces; b) no menos de dos veces.

Respuesta a) $P = P_6(0) + P_6(1) = \frac{7}{64}$;

b) $Q = 1 - [P_6(0) + P_6(1)] = \frac{57}{64}$.

6. La probabilidad de hacer blanco en un solo disparo de arma es $p = 0{,}9$. La probabilidad de batir el blanco en k impactos ($k \geqslant 1$) es igual a $1 - q^k$. Hallar la probabilidad de que se haga blanco si se efectuaron 2 disparos.

Advertencia. Utilícese las fórmulas de Bernoulli y de la probabilidad completa.

Respuesta 0,9639.

7. Hallar la probabilidad aproximada de que en 400 pruebas un suceso ocurrirá exactamente 104 veces si la probabilidad de su aparición en cada prueba es igual a 0,2.

Respuesta $P_{400}(104) = 0{,}0006$.

8. La probabilidad de que un tirador haga impacto en el blanco en un solo disparo es igual a 0,75. Hallar la probabilidad de que en 100 disparos se hará blanco: a) no menos de 70 y no más de 80 veces, b) no más de 70 veces.

Respuesta a) $P_{100}(70,\ 80) = 2\Phi(1{,}15) = 0{,}7498$;
b) $P_{100}(0;\ 70) = -\Phi(1{,}15) + 0{,}5 = 0{,}1251$.

9. La probabilidad de que ocurra un suceso en cada una de las 10 000 pruebas independientes es $p = 0{,}75$. Hallar la probabilidad de que la frecuencia relativa de aparición del suceso se desvía de su probabilidad no más de 0,001 en valor absoluto.

Respuesta $P = 2\Phi(0{,}23) = 0{,}182$.

10. La probabilidad de que ocurra un suceso en cada una de las pruebas independientes es igual a 0,2. Hallar qué desviación de la frecuencia relativa de aparición de un suceso respecto de su probabilidad se puede esperar con la probabilidad 0,9128 en 5000 experimentos.

Respuesta $\varepsilon = 0{,}00967$.

11. ¿Cuántas veces hay que arrojar una moneda para que con la probabilidad 0,6 se pueda esperar que la desviación de la frecuencia relativa de aparición de la cara respesto de la probabilidad $p = 0{,}5$ resulte en valor absoluto no mayor de 0,01?

Respuesta $n = 1764$.

Parte segunda

Magnitudes aleatorias

Capítulo sexto

TIPOS DE MAGNITUDES ALEATORIAS. DETERMINACION DE UNA MAGNITUD ALEATORIA DISCRETA

§ 1. Magnitud aleatoria

En la primera parte se citaron los sucesos constituidos por la aparición de uno u otro *número*. Por ejemplo, al tirar un dado pueden aparecer los números 1, 2, 3, 4, 5 y 6. No se puede determinar de antemano el número de puntos caídos, puesto que ello depende de muchas causas fortuitas que no son posibles de tomar en consideración íntegramente. En este sentido, el número de puntos es una magnitud aleatoria; los números 1, 2, 3, 4, 5 y 6 son los *valores posibles* de esta magnitud.

Se llama *aleatoria* la magnitud que como resultado de un experimento toma uno y solamente un valor posible, de antemano desconocido y dependiente de causas fortuitas que previamente no se pueden tener en cuenta.

Ejemplo 1. El número de niños nacidos entre cien recién nacidos es una magnitud aleatoria que tiene los siguientes valores posibles: 0, 1, 2, . . ., 100.

Ejemplo 2. La distancia que recorre un proyectil al disparar con un cañón es una magnitud aleatoria. En efecto, la distancia depende no sólo de la puesta de alza, sino también de muchas otras causas (fuerza y dirección del viento, temperatura, etc.) que no pueden ser consideradas enteramente.

Los valores posibles de esta magnitud corresponden a un cierto intervalo (a, b).

En adelante las magnitudes aleatorias las designaremos por las letras mayúsculas X, Y, Z, y sus valores posibles, respectivamente, por las minúsculas x, y, z. Por ejemplo, si la magnitud aleatoria X tiene tres valores posibles, éstas las designaremos así: x_1, x_2, x_3.

§ 2. Magnitudes aleatorias discretas y continuas

Volvamos a los ejemplos expuestos antes. En el primero de ellos la magnitud aleatoria X podría tomar uno de los siguientes valores posibles: 0, 1, 2, . . ., 100. Estos valores están separados entre sí por intervalos, en los que no hay valores posibles de X. En consecuencia, en este ejemplo la magnitud aleatoria toma valores posibles individuales aislados.

En el segundo ejemplo la magnitud aleatoria podría tomar cualquiera de los valores del intervalo (a, b). Aquí no se puede separar un valor posible de otro por un intervalo que no contenga valores posibles de la magnitud aleatoria.

De lo expuesto se deduce la conveniencia de distinguir las magnitudes aleatorias que toman sólo valores individuales, aislados, y las magnitudes aleatorias, cuyos valores posibles llenan enteramente cierto intervalo.

Se llama *discreta* (*discontinua*) la magnitud aleatoria que toma valores posibles individuales aislados con probabilidades determinadas. El número de valores posibles de una magnitud aleatoria discreta puede ser finito o infinito.

Se llama *continua* la magnitud aleatoria que puede tomar todos los valores de un cierto intervalo finito o infinito. Evidentemente, el número de valores posibles de una magnitud aleatoria continua es infinito.

Nota. Esta definición de la magnitud aleatoria continua no es exacta. Más adelante daremos una definición más rigurosa.

§ 3. Ley de distribución de probabilidades de una magnitud aleatoria discreta

A primera vista puede parecer que para prefijar una magnitud aleatoria discreta es suficiente enumerar todos sus valores posibles. En realidad esto no es así: las magnitudes aleatorias pueden tener enumeraciones *idénticas* de valores posibles, y sus probabilidades son *distintas*. Por eso, para prefijar una magnitud aleatoria discreta no es suficiente enumerar todos sus valores posibles, sino que hay que indicar también sus probabilidades.

Se llama *ley de distribución de una magnitud aleatoria discreta* la correspondencia entre los valores posibles y sus probabilidades; ésta puede prefijarse tabulada, analíticamente (en forma de fórmula) y gráficamente.

Cuando la ley de distribución de una magnitud aleatoria discreta se da tabulada, la primera línea de la tabla contiene los valores posibles y la segunda, sus probabilidades:

$$\begin{matrix} X & x_1 & x_2 \ldots & x_n \\ p & p_1 & p_2 \ldots & p_n. \end{matrix}$$

Tomando en consideración que en un experimento la magnitud aleatoria toma uno y solamente un valor posible, deducimos que los sucesos $X = x_1$, $X = x_2$, . . ., $X = x_n$ forman un grupo completo; por lo tanto, la suma de las probabilidades de estos sucesos, es decir, la suma de las probabilidades de la segunda línea es igual a la unidad:

$$p_1 + p_2 + \ldots + p_n = 1.$$

Ejemplo. En una lotería se han emitido 100 billetes. Se sortea un premio de 50 rublos y diez premios de 1 rublo cada uno. Hallar la ley de distribución de la magnitud aleatoria X, es decir, el valor del premio posible para el poseedor de un billete de lotería.

SOLUCION. Escribimos los valores posibles de X:

$$x_1 = 50, \qquad x_2 = 1, \qquad x_3 = 0.$$

Las probabilidades de estos valores posibles son:

$$p_1 = 0{,}01, \qquad p_2 = 0{,}1, \qquad p_3 = 1 - (p_1 + p_2) = 0{,}89.$$

Escribimos la ley de distribución buscada:

$$\begin{matrix} X & 50 & 10 & 0 \\ p & 0{,}01 & 0{,}1 & 0{,}89 \end{matrix}$$

Verificación: $0{,}01 + 01, + 0{.}89 = 1$.

Para claridad la ley de distribución de la magnitud aleatoria discreta puede representarse también gráficamente, para lo cual en el sistema de coordenadas rectangulares se marcan los puntos (x_i, p_i) y luego se unen por segmentos de rectas. La figura obtenida se llama *polígono de distribución*.

§ 4. Distribución binominal

Supongamos que se realizan n experimentos independientes, en cada uno de los cuales el suceso A puede ocurrir, o no ocurrir. La probabilidad de que ocurra el suceso en todas las pruebas es constante e igual a p (por lo tanto, la probabilidad de que no ocurra es $q = 1 - p$). Consideremos como mag-

nitud aleatoria discreta X el número de apariciones del suceso A en estas pruebas.

Planteémonos el problema: hallar la ley de distribución de la magnitud X. Para su resolución hay que determinar los valores posibles de X y sus probabilidades.

Evidentemente, el suceso A en n pruebas puede no aparecer, o aparece 1 vez, o 2 veces, . . ., o n veces. Por consiguiente, los valores posibles de X son:

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 2, \ldots, x_{n+1} = n.$$

Queda por hallar las probabilidades de estos valores posibles, para lo cual es suficiente utilizar la fórmula de Bernoulli:

$$P_n(k) = C_n^k p^k q^{n-k}. \qquad (*)$$

donde $k = 0, 1, 2, \ldots, n$.

La fórmula (*) es precisamente la expresión analítica de la ley de distribución buscada.

Se llama *binominal* la distribución de la probabilidad determinable por la fórmula de Bernoulli.

La ley se llama «binominal» puesto que el segundo miembro de la igualdad (*) se puede considerar como término general de la descomposición del binomio de Newton

$$(p+q)^n = C_n^n p^n + C_n^{n-1} p^{n-1} q + \ldots + C_n^k p^k q^{n-k} + \ldots + C_n^0 q^n.$$

De este modo, el primer término de la descomposición p^n determina la probabilidad de que el suceso examinado ocurra n veces en n pruebas independientes; el segundo término $np^{n-1}q$ determina la probabilidad de que el suceso ocurra $n-1$ veces; . . . ; el último término q^n determina la probabilidad de que el suceso no ocurra ni una sola vez.

Escribimos la ley binominal en forma de tabla:

X	n	$n-1$	...	k	...	0
P	p^n	$np^{n-1}q$	...	$C_n^k p^k q^{n-k}$	...	q^n.

Ejemplo. Una moneda se arroja dos veces. Escribir en forma de tabla la ley de distribución de la magnitud aleatoria X, es decir, el número de caídas de cara.

SOLUCION. La probabilidad de que aparezca la cara cada vez que se arroja la moneda es $p = \frac{1}{2}$, por lo tanto, la probabilidad de que no aparezca la cara es $q = 1 - \frac{1}{2} = \frac{1}{2}$.

Si la moneda se arroja dos veces la cara puede aparecer 2 veces, o 1 vez, o no aparecer. En consecuencia, los valores posibles de X son:

$$x_1 = 2, \qquad x_2 = 1, \qquad x_3 = 0.$$

Por la fórmula de Bernoulli hallamos las probabilidades de estos valores posibles:

$$P_2(2) = C_2^2 p^2 = \left(\frac{1}{2}\right)^2 = 0{,}25,$$

$$P_2(1) = C_2^1 pq = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = 0{,}5,$$

$$P_2(0) = C_2^0 q^2 = \left(\frac{1}{2}\right)^2 = 0{,}25.$$

Escribimos la ley de distribución buscada:

X	2	1	0
p	0,25	0,5	0,25.

Verificación: $0{,}25 + 0{,}5 + 0{,}25 = 1$.

§ 5. Distribución de Poisson

Supongamos que se realizan n experimentos independientes, en cada uno de los cuales la probabilidad de que aparezca el suceso A es igual a p. Para determinar la probabilidad de apariciones k del suceso en estas pruebas se utiliza la fórmula de Bernoulli. Si n es grande, se utiliza la forma asintótica de Laplace. Sin embargo, esta fórmula no es útil si la probabilidad del suceso es pequeña ($p \leqslant 0{,}1$). En estos casos (n es grande y p pequeño) se emplea a la fórmula asintótica de Poisson.

De esta manera, nos planteamos hallar la probabilidad de que para un número muy grande de pruebas, en cada una de las cuales la probabilidad del suceso es muy pequeña, el suceso ocurrirá exactamente k veces.

Hagamos una importante admisión: el producto np conserva un valor constante, precisamente $np = \lambda$. Como es de esperar de lo ulterior (cap. VII, § 5) esto significa que el promedio de apariciones del suceso en diferentes series de experimentos, es decir, para distintos valores de n, permanece invariable.

Utilizamos la fórmula de Bernoulli para calcular la probabilidad que nos interesa:

$$P_n(k)=\frac{n(n-1)(n-2)\ldots[n-(k-1)]}{k!}p^k(1-p)^{n-k}.$$

Puesto que $pn=\lambda$, $p=\frac{\lambda}{n}$. Por lo tanto,

$$P_n(k)=\frac{n(n-1)(n-2)\ldots[n-(k-1)]}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

Tomando en consideración que n tiene un valor muy grande, en lugar de $P_n(k)$ hallamos $\lim\limits_{n\to\infty}P_n(k)$. En este caso se hallará sólo el valor aproximado de la probabilidad buscada: aunque n es grande, pero finito y al hallar el límite hacemos tender n a infinito.

Así pues,

$$P_n(k)\simeq\lim_{n\to\infty}\frac{n(n-1)(n-2)\ldots[n-(k-1)]}{k!}\cdot\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}=$$

$$=\frac{\lambda^k}{k!}\lim_{n\to\infty}\left[1\cdot\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\right.$$

$$\left.\ldots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}\right]=$$

$$=\frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n\cdot\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=\frac{\lambda^k}{k!}\cdot e^{-\lambda}\cdot 1.$$

De este modo (para sencillez de la inscripción se ha omitido el signo de igualdad aproximada),

$$P_n(k)=\frac{\lambda^k e^{-\lambda}}{k!}.$$

Esta fórmula expresa la ley de distribución de Poisson de las probabilidades de sucesos raros (p pequeño) de masas (n grande).

Nota. Existen tablas especiales, con cuyo empleo se puede hallar $P_n(k)$ conociendo k y λ.

Ejemplo. Una fábrica envió al depósito 5000 piezas de buena calidad. La probabilidad de que durante el transporte una pieza se deteriora, es igual a 0,0002. Hallar la probabilidad de que lleguen al depósito 3 piezas inservibles.

SOLUCION. Por los datos del problema $n=5000$, $p=0{,}0002$, $k=3$. Hallamos λ:

$$\lambda=np=5000\cdot 0{,}0002=1.$$

La probabilidad buscada por la fórmula de Poisson es aproximadamente igual a

$$P_{5000}(3) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{e^{-1}}{3!} = \frac{1}{6e} \simeq 0{,}06.$$

§ 6. Flujo elemental de sucesos

Veamos los sucesos que ocurren en instantes aleatorios de tiempo.

Se llama *flujo de sucesos* la sucesión de sucesos que ocurren en instantes aleatorios de tiempo. Como ejemplos de flujos podemos citar: las llamadas a una central telefónica automática (CTA), a un centro de servicio médico de urgencia; la llegada de aviones al aeropuerto, de los clientes a un taller de servicio cotidiano, la sucesión de fallos de elementos, etc.

Entre las propiedades, propias de los flujos, distinguimos las de calidad de estacionario, de ausencia de efecto posterior y de calidad de ordinario.

La propiedad de calidad de estacionario se determina con que la probabilidad de que ocurran k sucesos en cualquier intervalo de tiempo depende solamente del número k y de la duración t del intervalo y no depende del comienzo de su lectura; en este caso, los distintos intervalos de tiempo se suponen no intersecables. Por ejemplo, las probabilidades de que ocurran k sucesos en los intervalos de tiempo (1; 7), (10; 16), (T; $T + 6$) de igual duración $t = 6$ unidades de tiempo, son iguales entre sí.

Así, *si un flujo posee la propiedad de calidad de estacionario, la probabilidad de que ocurran k sucesos en un intervalo de tiempo de duración t es una función que depende solamente de k y t.*

La propiedad de «ausencia de efecto posterior» se caracteriza con que la probabilidad de que ocurran k sucesos en cualquier intervalo de tiempo no depende de que hayan ocurrido o no los sucesos en los instantes de tiempo que preceden al comienzo del intervalo considerado. En otras palabras, la probabilidad condicional de que ocurran k sucesos en cualquier intervalo de tiempo, calculada para todas las suposiciones de lo que ocurrió hasta el comienzo del intervalo considerado (cuántos sucesos ocurrieron, en qué sucesión), es igual a la probabilidad incondicional. De este modo, la pre-

historia del flujo no se manifiesta sobre las probabilidades de que sucesos ocurran en un futuro próximo.

Así, *si un flujo posee la propiedad de ausencia de efecto posterior, hay la independencia mutua de apariciones de un número u otro de sucesos en intervalos de tiempo no intersecables.*

La propiedad de calidad de ordinario se caracteriza con que la aparición de dos y más sucesos en un intervalo de tiempo pequeño, prácticamente es imposible. En otras palabras, la probabilidad de que ocurra más de un suceso en un pequeño intervalo de tiempo es despreciable en comparación con la probabilidad de que ocurra solamente un suceso.

Así, *si un flujo tiene la propiedad de calidad de ordinario, en un intervalo de tiempo infinitamente pequeño puede ocurrir no más de un suceso.*

El flujo de sucesos que posee las propiedades de calidad de estacionario, de ausencia de efecto posterior y de calidad de ordinario se llama *simple o elemental (de Poisson).*

Nota. Generalmente es difícil establecer en la práctica si el flujo tiene las propiedades antes enumeradas. Por eso, se han hallado otras condiciones, cumpliéndose las cuales el flujo puede considerarse elemental, o próximo al elemental. En particular, se ha establecido que *si un flujo es la suma de un número muy grande de flujos estacionarios independientes, la influencia de cada uno de los cuales sobre la suma* (flujo total) *es despreciable, el flujo total* (en condición de su calidad de ordinario) *es próximo al elemental.*

El promedio de sucesos que ocurren en la unidad de tiempo se llama *intensidad del flujo* λ.

Se puede demostrar que si la intensidad constante del flujo es conocida, la probabilidad de que ocurran k sucesos de un flujo elemental por un intervalo de tiempo t se determina por la fórmula de Poisson

$$P_t(k) = \frac{(\lambda t)^k \cdot e^{-\lambda t}}{k!} .$$

Esta fórmula refleja todas las propiedades del flujo elemental.

En efecto, de la fórmula se aprecia que la probabilidad de que ocurran k sucesos en el tiempo t, para una intensidad dada, es una función de k y t, lo que determina la propiedad de carácter estacionario.

La fórmula no utiliza la información sobre la aparición de los sucesos hasta el comienzo del intervalo considerado, lo que caracteriza la propiedad de ausencia de efecto posterior.

Demostremos que la fórmula refleja la propiedad de calidad de ordinario. Poniendo $k = 0$ y $k = 1$, hallamos respectivamente las probabilidades de que no ocurran los sucesos y de que ocurra un solo suceso:

$$P_t(0) = e^{-\lambda t}, \quad P_t(1) = \lambda t e^{-\lambda t}.$$

Por lo tanto, la probabilidad de que ocurra más de un suceso es

$$P_t(k > 1) = 1 - [P_t(0) + P_t(1)] = 1 - [e^{-\lambda t} + \lambda t e^{-\lambda t}].$$

Descomponiendo

$$e^{-\lambda t} = 1 - \lambda t + \frac{(\lambda t)^2}{2!} - \ldots,$$

después de transformaciones simples, obtenemos

$$P_t(k > 1) = \frac{(\lambda t)^2}{2} + \ldots .$$

Comparando $P_t(1)$ y $P_t(k > 1)$, deducimos que para pequeños valores de t la probabilidad de que ocurra más de un suceso es despreciable en comparación con la probabilidad de que ocurra un suceso, lo que determina la propiedad de calidad de ordinario.

De esta manera, la fórmula de Poisson puede considerarse un modelo matemático de un flujo elemental de sucesos.

Ejemplo. El promedio de llamadas recibidas por una CTA en un minuto, es igual a dos. Hallar las probabilidades de que en 5 minutos se reciban: a) 2 llamadas, b) menos de dos llamadas, c) no menos de dos llamadas. El flujo de llamadas se supone elemental.

SOLUCION. Por los datos del problema $\lambda = 2$, $t = 5$, $k = 2$. Utilizamos la fórmula de Poisson

$$P_t(k) = \frac{(\lambda t)^k \cdot e^{-\lambda t}}{k!}.$$

a) La probabilidad buscada de que en 5 minutos se reciban 2 llamadas

$$P_5(2) = \frac{10^2 \cdot e^{-10}}{2!} = \frac{100 \cdot 0{,}000045}{2} = 0{,}00225.$$

Este suceso prácticamente es imposible.

b) Los sucesos «no se recibió ninguna llamada» y «se recibió una llamada» son mutuamente excluyentes, por eso la probabilidad buscada de que en 5 minutos se reciba menos de dos llamadas, por el teorema de adición, es

$$P_5(k<2)=P_5(0)+P_5(1)=e^{-10}+\frac{10\cdot e^{-10}}{1!}=0{,}000495.$$

Este suceso prácticamente es imposible.

c) Los sucesos «se recibió menos de dos llamadas» y «se recibieron no más de dos llamadas» son opuestos, por eso la probabilidad buscada de que en 5 minutos se reciban no menos de dos llamadas es

$$P_5(k\geqslant 2)=1-P_5(k<2)=1-0{,}000495=0{,}999505.$$

Este suceso es prácticamente cierto.

Problemas

1. Los valores posibles de una magnitud aleatoria son: $x_1=2$, $x_2=5$, $x_3=8$. Se conocen las probabilidades de los dos primeros valores posibles: $p_1=0{,}4$, $p_2=0{,}15$. Hallar la probabilidad de x_3.

Respuesta $p_3=0{,}45$.

2. Un dado se ha tirado 3 veces. Escribir la ley de distribución del número de apariciones del seis.

Respuesta

X	3	2	1	0
p	$\frac{1}{216}$	$\frac{15}{216}$	$\frac{75}{216}$	$\frac{125}{216}$

3. Componer la ley de distribución de las probabilidades del número de apariciones de un suceso A en tres pruebas independientes, si la probabilidad de que aparezca el suceso en cada una de las pruebas es igual a 0,6.

Respuesta

k	0	1	2	3
p	0,064	0,288	0,432	0,216

4. Una hiladora atiende 1000 husos. La probabilidad de que se corte el hilo en un huso durante un minuto es igual a 0,004. Hallar la probabilidad de que en un minuto se produzca el corte en cinco husos.

Respuesta $P_{1000}(5)=0{,}1562$.

5. Hallar el promedio de errores en una página manuscrita, si la probabilidad de que la página manuscrita contenga por lo menos un error, es igual a 0,95. Se supone que el número de errores está distribuido según la ley de Poisson.

Advertencia: el problema se reduce a buscar el parámetro λ de la ecuación $e^{-\lambda}=0{,}05$.

Respuesta 3.

6. El conmutador de una institución atiende 100 abonados. La probabilidad de que durante un minuto un abonado llame al conmutador, es igual a 0,02. ¿Cuál de los dos sucesos es más probable: en un minuto llaman 3 abonados; llaman 4 abonados?

Respuesta $P_{100}(3) = 0{,}18$; $P_{100}(4) = 0{,}09$.

7. Un original de 1000 páginas escritas a máquina contiene 1000 errores. Hallar la probabilidad de que una página tomada al azar contiene: a) por lo menos un error, b) exactamente 2 errores, c) no menos de dos errores. Se supone que el número de errores está distribuido por la ley de Poisson.

Respuesta a) $P = 1 - e^{-1} = 0{,}6321$, b) $P_{1000}(2) = 0{,}18395$, c) $P = 0{,}2642$.

8. El promedio de llamadas recibidas por una CTA en un minuto es igual a cinco. Hallar la probabilidad de que en 2 minutos se reciben: a) 2 llamadas, b) menos de dos llamadas, c) no menos de dos llamadas.
Advertencia: $e^{-10} = 0{,}000045$.

Respuesta a) 0,00225, b) 0,000495, c) 0,999505.

Capítulo séptimo

ESPERANZA MATEMATICA DE UNA MAGNITUD ALEATORIA DISCRETA

§ 1. Características numéricas de magnitudes aleatorias discretas

Ya sabemos que la ley de distribución caracteriza completamente la magnitud aleatoria. Sin embargo, de ordinario la ley de distribución no se conoce y hay que limitarse a menos datos. A veces, incluso conviene utilizar números que describan la magnitud aleatoria en total; estos números se llaman *características numéricas de una magnitud aleatoria*. Entre las características numéricas más importantes cabe mencionar la esperanza matemática.

Como se demostrará más adelante, la esperanza matemática es aproximadamente igual al valor medio de la magnitud aleatoria.

Para la resolución de muchos problemas es suficiente conocer la esperanza matemática. Por ejemplo, si se sabe que la esperanza matemática del número de puntos marcados por

el primer tirador es mayor que el de los marcados por el segundo, el primer tirador marcará un promedio mayor de puntos que el segundo y, por lo tanto, dispara mejor que el segundo.

A pesar de que la esperanza matemática da una cantidad bastante menor de datos sobre la magnitud aleatoria que la ley de su distribución, para resolver los problemas, como el expuesto y otros más, resulta suficiente conocer la esperanza matemática.

§ 2. Esperanza matemática de una magnitud aleatoria discreta

Se llama *esperanza matemática* de una magnitud aleatoria la suma de los productos de todos sus valores posibles por sus probabilidades.

Supongamos que una magnitud aleatoria X puede tomar solamente los valores $x_1, x_2, \ldots, x_n$, cuyas probabilidades respectivamente son $p_1, p_2, \ldots, p_n$. En ese caso, la esperanza matemática $M(X)$ de la magnitud aleatoria X se determina por la igualdad

$$M(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n.$$

Nota. De la definición se deduce que la esperanza matemática de una magnitud aleatoria discreta es una magnitud *no aleatoria* (constante). Recomendamos no olvidar esta afirmación, puesto que más adelante se utiliza con frecuencia. En adelante el lector sabrá que la esperanza matemática de una magnitud aleatoria continua también es una magnitud constante.

Ejemplo 1. Hallar la esperanza matemática de una magnitud aleatoria X, conociendo su ley de distribución:

X	3	5	2
p	0,1	0,6	0,3.

SOLUCION. La esperanza matemática buscada es igual a la suma de los productos de todos los valores posibles de la magnitud aleatoria por sus probabilidades:

$$M(X) = 3 \cdot 0{,}1 + 5 \cdot 0{,}6 + 2 \cdot 0{,}3 = 3{,}9.$$

Ejemplo 2. Hallar la esperanza matemática del número de apariciones del suceso A en una prueba, si la probabilidad del suceso A es igual a p.

SOLUCION. La magnitud aleatoria X, o sea, el número de apariciones del suceso A en una prueba puede tomar solamente dos valores: $x_1 = 1$ (el suceso A ocurrió) con la probabilidad p y $x_2 = 0$ (el suceso A no ocurrió) con la probabilidad $q = 1 - p$. La esperanza matemática buscada es

$$M(X) = 1 \cdot p + 0 \cdot q = p.$$

De suerte que *la esperanza matemática del número de apariciones de un suceso en una prueba es igual a la probabilidad de ese suceso*. Este resultado se utilizará más adelante.

§ 3. Sentido probabilístico de la esperanza matemática

Supongamos que se han realizado n pruebas, en las cuales la magnitud aleatoria X ha tomado m_1 vez el valor x_1, m_2 vez el valor x_2, . . ., m_k vez el valor x_k; además, $m_1 + m_2 + \ldots + m_k = n$. En ese caso, la suma de todos los valores tomados por X, es igual a

$$x_1m_1 + x_2m_2 + \ldots + x_km_k.$$

Hallamos la media aritmética $\overline{X}$ de todos los valores tomados por la magnitud aleatoria, para lo cual dividimos la suma obtenida por el número total de pruebas:

$$\overline{X} = \frac{x_1m_1 + x_2m_2 + \ldots + x_km_k}{n},$$

o bien

$$\overline{X} = x_1 \cdot \frac{m_1}{n} + x_2 \cdot \frac{m_2}{n} + \ldots + x_k \frac{m_k}{n}. \qquad (*)$$

Cabe hacer notar que la relación $\frac{m_1}{n}$ es la frecuencia relativa W_1 del valor x_1, $\frac{m_2}{n}$ es la frecuencia relativa W_2 del valor x_2, etc., por lo cual la correlación (*) podemos escribirla así:

$$\overline{X} = x_1W_1 + x_2W_2 + \ldots + x_kW_k. \qquad (**)$$

Supongamos que el número de pruebas es bastante grande. Entonces la frecuencia relativa es aproximadamente igual a la probabilidad de aparición del suceso (esto se demostrará en el cap IX, § 6):

$$W_1 \simeq p_1, \quad W_2 \simeq p_2, \ \ldots, \ W_k \simeq p_k.$$

Sustituyendo en la correlación (**) las frecuencias relativas por las correspondientes probabilidades, obtenemos

$$\overline{X} \simeq x_1p_1 + x_2p_2 + \ldots + x_kp_k.$$

El segundo miembro de esta igualdad aproximada es $M(X)$. De tal manera,

$$\overline{X} \simeq M(X).$$

El sentido probabilístico del resultado obtenido es: *la esperanza matemática es aproximadamente igual* (tanto más exacto, cuanto mayor es el número de pruebas) *a la media aritmética de los valores observados de la magnitud aleatoria.*

Nota 1. Se comprende fácilmente que la esperanza matemática es mayor que el mínimo y menor que el máximo de los valores posibles. En otras palabras, en el eje numérico los valores posibles están dispuestos a la izquierda y a la derecha de la esperanza matemática. En este sentido, la esperanza matemática caracteriza *la situación de la distribución* y, por eso, de ordinario se llama *centro de distribución.*

Este término ha sido adoptado de mecánica: si las masas $p_1, p_2, \ldots, p_n$ están situadas en los puntos de abscisas $x_1, x_2, \ldots, x_n$ además $\sum p_i = 1$, tendremos que la abscisa del centro de gravedad

$$x_c = \frac{\sum x_i p_i}{p_i}.$$

Teniendo en cuenta que $\sum x_i p_i = M(X)$ y $\sum p_i = 1$, obtenemos

$$M(X) = x_c.$$

De suerte que la esperanza matemática es la abscisa del centro de gravedad de un sistema de puntos materiales, cuyas abscisas son iguales a los valores posibles de la magnitud aleatoria y sus masas, a sus probabilidades.

Nota 2. El origen del término «esperanza matemática» está vinculado con el período inicial de surgimiento de la teoría de las probabilidades (siglos XVI — XVII), cuando el campo de su aplicación se limitaba a los juegos de azar. Al jugador le interesaba el valor medio del premio esperado o en otras palabras, la esperanza matemática del premio.

§ 4. Propiedades de la esperanza matemática

Propiedad 1. *La esperanza matemática de una magnitud constante es igual a la misma constante:*

$$M(C) = C.$$

DEMOSTRACION. La constante C la consideraremos como una magnitud aleatoria discreta que tiene sólo un valor posible C y lo toma con la probabilidad $p = 1$. Por lo tanto,

$$M(C) = C \cdot 1 = C.$$

Nota 1. Definimos *el producto de la magnitud constante C por la magnitud aleatoria discreta X* como una magnitud aleatoria discreta CX, cuyos valores posibles son iguales a los productos de la constante C por los valores posibles de X; las probabilidades de los valores posibles de CX son iguales a las probabilidades de los correspondientes valores posibles de X. Por ejemplo, si la probabilidad del valor posible x_1 es p_1, la probabilidad de que la magnitud CX toma el valor Cx_1 también es igual a p_1.

Propiedad 2. *Un factor constante se puede sacar fuera del signo de esperanza matemática:*

$$M(CX) = C \cdot M(X).$$

DEMOSTRACION. Supongamos que la magnitud aleatoria X está dada por la ley de distribución de las probabilidades:

$$\begin{array}{ccccc} X & x_1 & x_2 & \ldots & x_n \\ p & p_1 & p_2 & \ldots & p_n. \end{array}$$

Tomando en consideración la nota 1, escribimos la ley de distribución de la magnitud aleatoria CX:

$$\begin{array}{ccccc} CX & Cx_1 & Cx_2 & \ldots & Cx_n \\ p & p_1 & p_2 & \ldots & p_n. \end{array}$$

La esperanza matemática de la magnitud aleatoria CX es

$$M(CX) = Cx_1p_1 + Cx_2p_2 + \ldots + Cx_np_n = $$
$$= C(x_1p_1 + x_2p_2 + \ldots + x_np_n) = CM(X).$$

Así,

$$M(CX) = CM(X).$$

Nota 2. Antes de pasar a la siguiente propiedad, cabe señalar que dos magnitudes aleatorias se llaman *independientes* cuando la ley de distribución de una de ellas no depende de qué valores posibles ha tomado la otra magnitud. En caso contrario, las magnitudes aleatorias son *dependientes*. Varias magnitudes aleatorias se llaman *mutuamente independientes* si la ley de distribución de cualquier número de ellas no dependen de qué valores posibles tomaron las restantes magnitudes.

Nota 3. Determinamos *el producto de las magnitudes aleatorias independientes* X e Y como una magnitud aleatoria XY, cuyos valores posibles son iguales a los productos de cada valor posible de X por cada valor posible de Y; las probabilidades de los valores posibles del producto XY son iguales a los productos de las probabilidades de los valores posibles de los factores. Por ejemplo, si la probabilidad del valor posible x_1 es igual a p_1, la probabilidad del valor posible y_1 es igual a g_1, la probabilidad del valor posible x_1y_1 es igual a p_1g_1.

Propiedad 3. *La esperanza matemática del producto de dos magnitudes aleatorias independientes es igual al producto de sus esperanzas matemáticas:*

$$M(XY) = M(X)\,M(Y).$$

DEMOCTRACION. Supongamos que las magnitudes aleatorias independientes X e Y están prefijadas por sus leyes de distribución de las probabilidades*:

XY	x_1	x_2		Y	y_1	g_2
P	p_1	p_2		g	g_1	y_2.

Componemos todos los valores que puede tomar la magnitud aleatoria XY, para lo cual multiplicamos todos los valores posibles de X por cada valor posible de Y; como resultado obtenemos: x_1y_1, x_2y_1, x_1y_2 y x_2y_2.

Teniendo en cuenta la nota 3, escribimos la «ley de distribución» del producto XY:

XY	x_1y_1	x_2y_1	x_1y_2	x_2y_2
P	p_1g_1	p_2g_1	p_1g_2	p_2g_2.

La esperanza matemática es igual a la suma de los productos de todos los valores posibles por sus probabilidades:
$M(XY) = x_1y_1\cdot p_1g_1 + x_2y_1\cdot p_2g_1 + x_1y_2\cdot p_1g_2 + x_2y_2\cdot p_2g_2$,
o bien

$$M(XY) = y_1g_1(x_1p_1 + x_2p_2) + y_2g_2(x_1p_1 + x_2p_2) =$$
$$= (x_1p_1 + x_2p_2)(y_1g_1 + y_2g_2) = M(X)\cdot M(Y).$$

De este modo, $M(XY) = M(X)\cdot M(Y)$.

---

* Nos hemos limitado a un número pequeño de valores posibles, para simplificar los cálculos. En el caso general, la demostración es análoga.

Corolario. *La esperanza matemática del producto de varias magnitudes aleatorias mutuamente independientes es igual al producto de sus esperanzas matemáticas.*

Por ejemplo, para tres magnitudes aleatorias tenemos:

$$M(XYZ) = M(XY \cdot Z) = M(XY)\, M(Z) =$$
$$= M(X)\, M(Y)\, M(Z).$$

Para un número arbitrario de magnitudes aleatorias la demostración se realiza por el método de inducción matemática.

Ejemplo 1. Las magnitudes aleatorias independientes X e Y están prefijadas por las siguientes leyes de distribución:

X	5	2	4		Y	7	9
p	0,6	0,1	0,3		p	0,8	0,2.

Hallar la esperanza matemática de la magnitud aleatoria XY.

SOLUCIÓN. Hallamos la esperanza matemática de cada una de las magnitudes dadas:

$$M(X) = 5 \cdot 0{,}6 + 2 \cdot 0{,}1 + 4 \cdot 0{,}3 = 4{,}4;$$
$$M(Y) = 7 \cdot 0{,}8 + 9 \cdot 0{,}2 = 7{,}4.$$

Puesto que las magnitudes aleatorias X e Y son independientes, la esperanza matemática buscada es igual a:

$$M(XY) = M(X)\, M(Y) = 4{,}4 \cdot 7{,}4 = 32{,}56.$$

Nota 4. Determinamos la *suma de las magnitudes aleatorias* X e Y como una magnitud aleatoria $X + Y$, cuyos valores posibles son iguales a las sumas de cada valor posible X con cada valor posible de Y; las probabilidades de los valores posibles de $X + Y$ para las magnitudes independientes X e Y son iguales a los productos de las probabilidades de los sumandos; para las magnitudes dependientes, a los productos de la probabilidad de un sumando por la probabilidad condicional del segundo.

La propiedad que se da a continuación se cumple tanto para las magnitudes aleatorias independientes, como para las dependientes.

Propiedad 4. *La esperanza matemática de la suma de dos magnitudes aleatorias es igual a la suma de las esperanzas matemáticas de los sumandos:*

$$M(X + Y) = M(X) + M(Y).$$

DEMOSTRACIÓN. Supongamos que las magnitudes aleatorias X e Y están prefijadas por las siguientes leyes de distribución*:

$$\begin{array}{ccc} X & x_1 & x_2 \\ p & p_1 & p_2 \end{array} \qquad \begin{array}{ccc} Y & y_1 & y_2 \\ g & g_1 & g_2. \end{array}$$

Componemos todos los valores posibles de la magnitud $X + Y$, para lo cual, a cada valor posible de X sumamos cada valor posible de Y; obtenemos $x_1 + y_1$, $x_1 + y_2$, $x_2 + y_1$ y $x_2 + y_2$. Las probabilidades de estos valores las designamos respectivamente por p_{11}, p_{12}, p_{21} y p_{22}.

La esperanza matemática de la magnitud $X + Y$ es igual a la suma de los productos de los valores posibles por sus probabilidades:

$$M(X + Y) = (x_1 + y_1)\, p_{11} + (x_1 + y_2)\, p_{12} + \\ + (x_2 + y_1)\, p_{21} + (x_2 + y_2)\, p_{22},$$

o bien

$$M(X + Y) = x_1 (p_{11} + p_{12}) + x_2 (p_{21} + p_{22}) + \\ + y_1 (p_{11} + p_{21}) + y_2 (p_{12} + p_{22}). \qquad (*)$$

Demostremos que $p_{11} + p_{12} = p_1$. El suceso constituido en que X toma el valor x_1 (la probabilidad de este suceso es igual a p_1), da lugar al suceso que consiste en que $X + Y$ toma el valor $x_1 + y_1$ o bien $x_1 + y_2$ (la probabilidad de este suceso por el teorema de la adición es igual a $p_{11} + p_{12}$) e inversamente. De aquí se deduce que $p_{11} + p_{12} = p_1$.

Análogamente se demuestran las igualdades

$$p_{21} + p_{22} = p_2,\ p_{11} + p_{21} = g_1 \text{ y } p_{12} + p_{22} = g_2.$$

Sustituyendo los primeros miembros de estas igualdades en la correlación (*), obtenemos:

$$M(X + Y) = (x_1 p_1 + x_2 p_2) + (y_1 g_1 + y_2 g_2),$$

o bien, finalmente

$$M(X + Y) = M(X) + M(Y).$$

Corolario. *La esperanza matemática de la suma de varias magnitudes aleatorias es igual a la suma de las esperanzas matemáticas de los sumandos.*

---

* Para simplificar la deducción nos limitamos sólo a dos valores posibles de cada una de las magnitudes. En el caso general, la demostración es análoga.

Por ejemplo, para tres magnitudes sumandos tendremos:

$$M(X+Y+Z) = M[(X+Y)+Z] =$$

$$= M(X+Y) + M(Z) = M(X) + M(Y) + M(Z).$$

Para un número arbitrario de magnitudes sumandos la demostración se realiza por el método de inducción matemática.

Ejemplo 1. Se efectúan 3 disparos con las probabilidades de hacer blanco, iguales a $p_1 = 0{,}4$; $p_2 = 0{,}3$ y $p_3 = 0{,}6$. Hallar la esperanza matemática del número total de impactos.

SOLUCIÓN. El número de impactos en el primer disparo es una magnitud aleatoria X_1, que puede tomar solamente dos valores: 1 (impacto) con probabilidad $p_1 = 0{,}4$ o bien 0 (fallo) con probabilidad $q = 1 - 0{,}4 = 0{,}6$.

La esperanza matemática del número de impactos en el primer disparo es igual a la probabilidad de impacto (véase el ejemplo de la pág. 80), es decir, $M(X_1) = 0{,}4$.

Análogamente hallamos la esperanza matemática del número de impactos en el segundo y tercero disparos:

$$M(X_2) = 0{,}3, \qquad M(X_3) = 0{,}6.$$

El número total de impactos también es una magnitud aleatoria compuesta de la suma de impactos en cada uno de los tres disparos:

$$X = X_1 + X_2 + X_3.$$

La esperanza matemática buscada la hallamos por el teorema de la esperanza matemática de la suma:

$$M(X) = M(X_1 + X_2 + X_3) = M(X_1) + M(X_2) +$$
$$+ M(X_3) = 0{,}4 + 0{,}3 + 0{,}6 = 1{,}3 \text{ (impactos)}.$$

Ejemplo 2. Hallar la esperanza matemática de la suma del número de puntos que pueden aparecer al tirar dos dados.

SOLUCIÓN. El número de puntos que puede aparecer en el primer dado, lo designamos por X y el del segundo, por Y. Los valores posibles de estas magnitudes son idénticos e iguales a 1, 2, 3, 4, 5 y 6; además, la probabilidad de cada uno de estos valores es igual a $\frac{1}{6}$.

Hallamos la esperanza matemática del número de puntos que pueden aparecer en el primer dado:

$$M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Es evidente que también $M(Y) = \frac{7}{2}$.

La esperanza matemática buscada es

$$M(X+Y) = M(X) + M(Y) = \frac{7}{2} + \frac{7}{2} = 7,$$

§ 5. Esperanza matemática del número de apariciones de un suceso en experimentos independientes

Supongamos que se realizan n pruebas independientes, en cada una de las cuales la probabilidad de que aparezca el suceso A es constante e igual a p. ¿A qué es igual el promedio de apariciones del suceso A en estas pruebas? La respuesta la da el siguiente teorema.

Teorema. *La esperanza matemática $M(X)$ del número de apariciones del suceso A en n pruebas independientes es igual al producto del número de pruebas por la probabilidad de que aparezca el suceso en cada prueba:*

$$M(X) = np.$$

DEMOSTRACION. Vamos a considerar como magnitud aleatoria X el número de apariciones del suceso A en n pruebas independientes.

Evidentemente, el número total X de apariciones del suceso A en esas pruebas se compone del número de apariciones del suceso en cada prueba. Por eso, si X_1 es el número de apariciones del suceso en la primera prueba, X_2 en la segunda, . . ., X_n en la n-esima, el número total de apariciones del suceso $X = X_1 + X_2 + \ldots X_n$.

Por la tercera propiedad de la esperanza matemática, tendremos

$$M(X) = M(X_1) + M(X_2) + \ldots + M(X_n). \qquad (*)$$

Cada uno de los sumandos del segundo miembro de la igualdad es la esperanza matemática del número de apariciones del suceso en una prueba: $M(X_1)$, en la primera, $M(X_2)$

en la segunda, etc. Puesto que la esperanza matemática del número de apariciones de un suceso en una prueba es igual a la probabilidad del suceso (§ 2, ejemplo 2), tendremos que $M(X_1) = M(X_2) = M(X_n) = p$. Poniendo en el segundo miembro de la igualdad (*) en lugar de cada sumando p, obtenemos

$$M(X) = np. \qquad (**)$$

Nota. Dado que la magnitud X está distribuida por la ley binomial, el teorema demostrado se puede formular también así: *la esperanza matemática de la distribución binomial* de parámetros n y p es igual al producto np.

Ejemplo. La probabilidad de hacer blanco al tirar desde un cañon es $p=0,6$. Hallar la esperanza matemática del número total de impactos, si se realizan 10 disparos.

SOLUCION. El impacto en cada disparo no depende de los resultados de los otros disparos, por lo cual los sucesos considerados son independientes y, por lo tanto, la esperanza matemática buscada es

$$M(X) = np = 10 \cdot 0,6 = 6 \text{ (impactos)}.$$

Problemas

1. Hallar la esperanza matemática de una magnitud aleatoria discreta, conociendo la ley de su distribución:

X	6	3	1
p	0,2	0,3	0,5

Respuesta 2,6.

2. Se efectuan 4 disparos con las probabilidades de impacto $p_1 = 0,6$, $p_2 = 0,4$, $p_3 = 0,5$ y $p_4 = 0,7$. Hallar la esperanza matemática del número total de impactos.

Respuesta 2,2 impactos.

3. Las magnitudes aleatorias independientes discretas están prefijadas por las leyes de distribución:

X	1	2
p	0,2	0,8

Y	0,5	1
p	0,3	0,7

Hallar la esperanza matemática del producto XY por dos métodos: 1) componiendo la ley de distribución de XY; 2) utilizando la propiedad 3.

Respuesta 1,53.

4. Las magnitudes aleatorias discretas X e Y están prefijadas por las leyes de distribución señaladas en el problema 3. Hallar la esperanza matemática de la suma $X + Y$ por dos métodos: 1) componiendo la ley de distribución de $X + Y$; 2) utilizando la propiedad 4.

Respuesta 2,65.

5. La probabilidad de que falle una pieza durante la prueba de fiabilidad es igual a 0,2. Hallar la esperanza matemática del número de piezas que fallan, si se verifican 10 piezas.

Respuesta 2 piezas.

6. Hallar la esperanza matemática del número de puntos que pueden salir al tirar simultáneamente dos dados.

Respuesta 12,25 puntos.

7. Hallar la esperanza matemática del número de billetes de lotería que pueden ser premiados, si se han adquirido 20 billetes, y la probabilidad de que sea premiado un billete es igual a 0,3.

Respuesta 6 billetes.

Capítulo octavo

DISPERSION DE UNA MAGNITUD ALEATORIA DISCRETA

§ 1. Utilidad de la introducción de la característica numérica de dispersión de una magnitud aleatoria

Es fácil predecir magnitudes aleatorias tales que tengan idénticas esperanzas matemáticas, pero distintos valores posibles.

Veamos, por ejemplo, las magnitudes aleatorias discretas X e Y prefijadas por las siguientes leyes de distribución:

X	$-0{,}01$	$0{,}01$	Y	-100	100
p	$0{,}5$	$0{,}5$	p	$0{,}5$	$0{,}5$.

Hallamos las esperanzas matemáticas de estas magnitudes:

$$M(X) = -0{,}01 \cdot 0{,}5 + 0{,}01 \cdot 0{,}5 = 0,$$

$$M(Y) = -100 \cdot 0{,}5 + 100 \cdot 0{,}5 = 0.$$

Aquí las esperanzas matemáticas de ambas magnitudes son idénticas, y los valores posibles son distintos; además,

X tiene valores posibles, próximos a la esperanza matemática, en tanto que Y, apartados de su esperanza matemática. En consecuencia, conociendo solamente la esperanza matemática de una magnitud aleatoria, no se puede predecir qué valores posibles tomará, ni tampoco cómo están dispersos alrededor de la esperanza matemática. En otras palabras, la esperanza matemática no caracteriza del todo la magnitud aleatoria.

Por esta causa, junto con la esperanza matemática se introducen también otras características numéricas. Así, por ejemplo, para estimar como están dispersos los valores posibles de una magnitud aleatoria alrededor de su esperanza matemática, se utiliza, en particular, una característica numérica, llamada dispersión.

Antes de pasar a la definición y a las propiedades de la dispersión, introducimos el concepto de desviación de una magnitud aleatoria respecto de su esperanza matemática.

§ 2. Desviación de una magnitud aleatoria de su esperanza matemática

Supongamos que X es una magnitud alaetoria y $M(X)$, su esperanza matemática. Consideremos como una nueva magnitud aleatoria la diferencia $X - M(X)$.

Se llama *desviación* la diferencia entre una magnitud aleatoria y su esperanza matemática.

Supongamos que se conoce la ley de distribución de X

$$\begin{array}{ccccc} X & x_1 & x_2 & \dots & x_n \\ p & p_1 & p_2 & \dots & p_n. \end{array}$$

Escribimos la ley de distribución de la desviación. Para que la desviación tome el valor $x_1 - M(X)$ es suficiente que la magnitud aleatoria tome el valor x_1. La probabilidad de este suceso es igual a p_1; por lo tanto, también la probabilidad de que la desviación tome el valor $x_1 - M(X)$, asimismo es igual a p_1. Análogamente se presenta para los demás valores posibles de la desviación.

Por consiguiente, la desviación tiene la siguiente ley de distribución

$$\begin{array}{ccccc} X - M(X) & x_1 - M(X) & x_2 - M(X) & \dots & x_n - M(X) \\ p & p_1 & p_2 & \dots & p_n. \end{array}$$

Damos una importante propiedad de la desviación que será utilizada más adelante.

Teorema. *La esperanza matemática de la desviación es igual a cero*:

$$M[X - M(X)] = 0.$$

DEMOSTRACION. Utilizando las propiedades de la esperanza matemática (la esperanza matemática de la diferencia es igual a la diferencia de las esperanzas matemáticas, la esperanza matemática de una constante es igual a misma constante) y tomando en consideración que $M(X)$ es una magnitud constante, tendremos:

$$M[X - M(X)] = M(X) - M[M(X)] = M(X) - \\ - M(X) = 0.$$

§ 3. Dispersión de una magnitud aleatoria discreta

De ordinario, en la práctica hay que estimar la dispersión de los valores posibles de una magnitud aleatoria alrededor de su valor medio. Por ejemplo, en artillería es importante saber con qué precisión caen los proyectiles cerca del blanco que debe ser batido.

A primera vista puede parecer que para estimar la dispersión lo más fácil es calcular todos los valores posibles de la desviación de una magnitud aleatoria y después hallar su valor medio. Sin embargo, este método no da resultado, ya que el valor medio de la desviación, es decir, $M[X - M(X)]$ para cualquier magnitud aleatoria es igual a cero. Esta propiedad ya fue demostrada en el párrafo anterior y se explica con que ciertas desviaciones posibles son positivas y otras negativas; debido a la anulación mutua el valor medio de la desviación es igual a cero.

Estas consideraciones denotan la conveniencia de sustituir las desviaciones posibles por sus valores absolutos o por sus cuadrados. Precisamente así se procede. En realidad, cuando las desviaciones posibles se sustituyen por sus valores absolutos, se debe operar con magnitudes absolutas, lo que, a veces, da lugar a serias dificultades. Por eso, frecuentemente se sigue otro método, es decir, se calcula el valor medio del cuadrado de la desviación que precisamente se llama dispersión.

Se llama *dispersión* de una magnitud aleatoria discreta la esperanza matemática del cuadrado de la desviación de la

magnitud aleatoria respesto de su esperanza matemática:

$$D(X) = M[X - M(X)]^2. = \sigma^2 = E[(x-\mu)^2]$$

Supongamos que una magnitud aleatoria está prefijada por la ley de distribución

X	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n.

En tal caso, el cuadrado de la desviación tiene la siguiente ley de distribución:

$[X-M(X)]^2$	$[x_1-M(X)]^2$	$[x_2-M(X)]^2$	...	$[x_n-M(X)]^2$
p	p_1	p_2	...	p_n.

Por definición la dispersión es igual a

$$D(X) = M[X - M(X)]^2 = [x_1 - M(X)]^2 \cdot p_1 + [x_2 - M(X)]^2 \times p_2 + \ldots + [x_n - M(X)]^2 \cdot p_n.$$

De esta manera, para hallar la dispersión es suficiente calcular la suma de los productos de los valores posibles del cuadrado de la desviación por sus probabilidades.

Nota. De la definición se deduce que la dispersión de una magnitud aleatoria discreta es una magnitud no aleatoria (constante). En adelante el lector sabrá que la dispersión de una magnitud aleatoria continua también es una magnitud constante.

Ejemplo. Hallar la dispersión de la magnitud aleatoria X, dada por la siguiente ley de distribución:

X	1	2	5
p	0,3	0,5	0,2.

SOLUCION. Hallamos la esperanza matemática

$$M(X) = 1 \cdot 0,3 + 2 \cdot 0,5 + 5 \cdot 0,2 = 2,3.$$

Hallamos todos los valores posibles del cuadrado de la desviación

$$[x_1 - M(X)]^2 = (1 - 2,3)^2 = 1,69;$$
$$[x_2 - M(X)]^2 = (2 - 2,3)^2 = 0,09;$$
$$[x_3 - M(X)]^2 = (5 - 2,3)^2 = 7,29.$$

Escribimos la ley de distribución del cuadrado de la desviación:

$[X-M(X)]^2$	1,69	0,09	7,29
p	0,3	0,5	0,2.

Por definición de dispersión

$$D(X) = 1{,}69\cdot 0{,}3 + 0{,}09\cdot 0{,}5 + 7{,}29\cdot 0{,}2 = 2{,}01.$$

Vemos que el cálculo basado en la definición de dispersión, resultó relativamente voluminoso. A continuación expondremos la fórmula que nos conduce bastante más rápidamente al resultado.

§ 4. Fórmula para el cálculo de la dispersión

Generalmente para el cálculo de la dispersión resulta conveniente utilizar el siguiente teorema.

Teorema. *La dispersión es igual a la diferencia entre la esperanza matemática del cuadrado de la magnitud aleatoria X y el cuadrado de su esperanza matemática:*

$$D(X) = M(X^2) - [M(X)]^2.$$

DEMOSTRACIÓN. La esperanza matemática $M(X)$ es una magnitud constante, por lo tanto, $2M(X)$ y $M^2(X)$ son también magnitudes constantes. Tomando en consideración esto y utilizando las propiedades de la esperanza matemática (el factor constante se puede extraer del signo de esperanza matemática; la esperanza matemática de la suma es igual a la suma de las esperanzas matemáticas de los sumandos), simplificamos la fórmula que expresa la definición de la dispersión:

$$D(X) = M[X - M(X)]^2 = M[X^2 - 2XM(X) +$$
$$+ M^2(X)] = M(X^2) - 2M(X)\cdot M(X) + M^2(X) =$$
$$= M(X^2) - 2M^2(X) + M^2(X) = M(X^2) - M^2(X).$$

De este modo,

$$D(X) = M(X^2) - [M(X)]^2.$$

Para recordar más fácilmente la fórmula se han introducido en su notación los corchetes.

Ejemplo 1. Hallar la dispersión de la magnitud aleatoria X prefijada por la siguiente ley de distribución:

X	2	3	5
p	0,1	0,6	0,3.

SOLUCION. Hallamos la esperanza matemática $M(X)$:

$$M(X) = 2\cdot 0{,}1 + 3\cdot 0{,}6 + 5\cdot 0{,}3 = 3{,}5.$$

Escribimos la ley de distribución de la magnitud aleatoria X^2:

X^2	4	9	25
p	0,1	0,6	0,3.

Hallamos la esperanza matemática $M(X^2)$:

$$M(X^2) = 4\cdot 0{,}1 + 9\cdot 0{,}6 + 25\cdot 0{,}3 = 13{,}3.$$

La dispersión buscada es

$$D(X) = M(X^2) - [M(X)]^2 = 13{,}3 - (3{,}5)^2 = 1{,}05.$$

Nota. Aparentemente, si X e Y tienen idénticos valores posibles e igual esperanza matemática, las dispersiones de estas magnitudes son iguales (en efecto, ¡los valores posibles de ambas magnitudes están igualmente dispersas alrededor de sus esperanzas matemáticas!). Empero, en el caso general, esto no es así. Resulta que los valores posibles idénticos de las magnitudes examinadas tienen, en general, distintas probabilidades, *en tanto que la magnitud de la dispersión se determina no sólo por los mismos valores posibles, sino también por sus probabilidades.* Por ejemplo, si las probabilidades «lejanas» respecto de la esperanza matemática de los valores posibles de X son mayores que las probabilidades de los mismos valores de Y y las probabilidades «próximas» de los valores de X son menores que las probabilidades de los mismos valores de Y, evidentemente, la dispersión de X es mayor que la dispersión de Y.

Veamos un ejemplo ilustrativo.

Ejemplo 2. Comparar las dispersiones de las magnitudes aleatorias prefijadas por las leyes de distribución:

X	-1	1	2	3
p	0,48	0,01	0,09	0,42

Y	-1	1	2	3
p	0,19	0,51	0,25	0,05

SOLUCION. Se aprecia fácilmente que

$$M(X) = M(Y) = 0{,}97;$$

$$D(X) \simeq 3{,}69, \qquad D(Y) \simeq 1{,}21.$$

Por lo tanto, los valores posibles y las esperanzas matemáticas de X e Y son idénticas, mientras que las dispersiones son distintas; además, $D(X) > D(Y)$.

Este resultado se podía haber previsto sin cálculos, observando solamente las leyes de distribución.

§ 5. Propiedades de la dispersion

Propiedad 1. *La dispersión de una magnitud constante C es igual a cero:*

$$D(C) = 0.$$

DEMOSTRACION. Por definición de dispersión tenemos que

$$D(C) = M\{[C - M(C)]^2\}.$$

Utilizando la primera propiedad de la esperanza matemática (la esperanza matemática de una constante es igual a la misma constante), obtenemos

$$D(C) = M[(C - C)^2] = M(0) = 0.$$

Así

$$D(C) = 0.$$

La propiedad se hace más clara si se tiene en cuenta que una magnitud constante conserva el mismo valor y, desde luego, no tiene dispersión.

Propiedad 2. *Un factor constante se puede sacar fuera del signo de dispersión, elevándola al cuadrado:*

$$D(CX) = C^2 D(X).$$

DEMOSTRACION. Por definición de dispersión tenemos que

$$D(CX) = M\{[CX - M(CX)]^2\}.$$

Utilizando la segunda propiedad de la esperanza matemática (el factor constante se puede sacar fuera del signo de esperanza matemática), obtenemos

$$D(CX) = M\{[CX - CM(X)]^2\} =$$

$$= M\{C^2[X - M(X)]^2\} = C^2 M\{[X - M(X)]^2\} = C^2 D(X).$$

De esta suerte

$$D(CX) = C^2 D(X).$$

La propiedad se hace más clara si se toma en consideración que cuando $|C| > 1$ la magnitud CX tiene valores posibles (en magnitud absoluta) mayores que la magnitud de X. De aquí se deduce que, estos valores dispersos alrededor de la esperanza matemática $M(CX)$ son mayores que los valores posibles de X alrededor de $M(X)$, es decir, $D(CX) > D(X)$. Por el contrario, si $0 < |C| < 1$, tendremos que $D(CX) < < D(X)$.

Propiedad 3. *La dispersión de la suma de dos magnitudes aleatorias independientes es igual a la suma de las dispersiones de estas magnitudes:*

$$D(X + Y) = D(X) + D(Y).$$

DEMOSTRACIÓN. Por la fórmula para el cálculo de la dispersión tenemos que

$$D(X + Y) = M[(X + Y)^2] - [M(X + Y)]^2.$$

Abriendo paréntesis y utilizando las propiedades de la esperanza matemática de la suma de varias magnitudes y del producto de dos magnitudes aleatorias independientes, obtenemos

$$D(X + Y) = M[X^2 + 2XY + Y^2] - [M(X) + M(Y)]^2 = M(X^2) + 2M(X)\cdot M(Y) + M(Y^2) - M^2(X) - 2M(X)\cdot M(Y) - M^2(Y) = \{M(X^2) - [M(X)]^2\} + \{M(Y^2) - [M(Y)]^2\} = D(X) + D(Y).$$

Así que

$$D(X + Y) = D(X) + D(Y).$$

Corolario 1. *La dispersión de la suma de varias magnitudes aleatorias mutuamente independientes es igual a la suma de las dispersiones de estas magnitudes.*

Por ejemplo, para tres sumandos tenemos que

$$D(X + Y + Z) = D[X + (Y + Z)] = D(X) + D(Y + Z) = D(X) + D(Y) + D(Z).$$

Para un número arbitrario de sumandos la demostración se realiza por el método de inducción matemática.

Corolario 2. *La dispersión de la suma de una magnitud constante y de una aleatoria es igual a la dispersión de la magnitud aleatoria:*

$$D(C + X) = D(X).$$

DEMOSTRACIÓN. Las magnitudes C y X son independientes, por eso, según la tercera propiedad

$$D(C + X) = D(C) + D(X).$$

En virtud de la primera propiedad $D(C) = 0$. Por consiguiente,

$$D(C + X) = D(X).$$

La propiedad se hace comprensible si se tiene en cuenta que las magnitudes X y $X + C$ se diferencian solamente por el comienzo de la lectura y por lo tanto, están dispersas igualmente alrededor de sus esperanzas matemáticas.

Propiedad 4. *La dispersión de la diferencia de dos magnitudes aleatorias independientes es igual a la suma de sus dispersiones*:

$$D(X - Y) = D(X) + D(Y).$$

DEMOSTRACION. En virtud de la tercera propiedad

$$D(X - Y) = D(X) + D(-Y).$$

Por la segunda propiedad

$$D(X - Y) = D(X) + (-1)^2 D(Y),$$

ó

$$D(X - Y) = D(X) + D(Y).$$

§ 6. Dispersión del número de apariciones de un suceso en experimentos independientes

Supongamos que se realizan n experimentos independientes, en cada uno de los cuales la probabilidad de que aparezca el suceso A es constante. ¿A qué es igual la dispersión del número de apariciones del suceso en estas pruebas? La respuesta la da el siguiente teorema.

Teorema. *La dispersión del número de apariciones de un suceso A en n pruebas independientes, en cada una de las cuales la probabilidad p de que ocurra el suceso es constante, es igual al producto del número de pruebas por las probabilidades de que aparezca y no aparezca el suceso en una prueba*:

$$D(X) = npq.$$

DEMOSTRACION. Consideremos la magnitud aleatoria X, o sea, el número de apariciones del suceso A en n pruebas independientes. Evidentemente, el número total de apariciones del suceso en estas pruebas es igual a la suma de las apariciones del suceso en pruebas individuales:

$$X = X_1 + X_2 + \ldots + X_n,$$

donde X_1 es el número de apariciones del suceso en la primera prueba; X_2, en la segunda, ... X_n, en la n-sima.

Las magnitudes $X_1, X_2, \ldots, X_n$ son mutuamente independientes, puesto que el resultado de cada prueba no depende de los resultados de las demás, por eso tenemos razón de utilizar el corolario 1 (§ 5):

$$D(X) = D(X_1) + D(X_2) + \ldots + D(X_n). \qquad (*)$$

Calculamos la dispersión de X_1 por la fórmula

$$D(X_1) = M(X_1^2) - [M(X_1)]^2. \qquad (**)$$

La magnitud X_1 es el número de apariciones del suceso A en la primera tentativa, por eso (cap. VII, § 2, ejemplo 2) $M(X_1) = p$.

Hallamos la esperanza matemática de la magnitud X_1^2 que puede tomar solamente dos valores, o sea, 1^2 con probabilidad p y 0^2 con probabilidad q:

$$M(X_1^2) = 1^2 \cdot p + 0^2 \cdot q = p.$$

Sustituyendo los resultados obtenidos en la correlación (**) tenemos que

$$D(X_1) = p - p^2 = p(1 - p) = pq.$$

Evidentemente, la dispersión de cada una de las magnitudes aleatorias restantes también es igual a p. Sustituyendo cada sumando del segundo miembro de (*) por p, finalmente obtenemos

$$D(X) = npq.$$

Nota. Puesto que la magnitud X está distribuida por la ley binomial, el teorema demostrado se puede formular también así: *la dispersión de la distribución binomial de parámetros n y p es igual al producto npq.*

Ejemplo. Se realizan 10 pruebas independientes en cada una de las cuales la probabilidad de que ocurra el suceso es igual a 0,6. Hallar la dispersión de la magnitud aleatoria X, o sea, del número de apariciones del suceso en estas pruebas.

SOLUCION. Por los datos del problema $n = 10$; $p = 0{,}6$. Evidentemente, la probabilidad de que no ocurra el suceso es

$$q = 1 - 0{,}6 = 0{,}4.$$

La dispersión buscada es

$$D(X) = npq = 10 \cdot 0{,}6 \cdot 0{,}4 = 2{,}4.$$

§ 7. Desviación cuadrática media

Para estimar la dispersión de los valores posibles de una magnitud aleatoria alrededor de su valor medio, además de la dispersión sirven también algunas otras características. Entre ellas podemos citar la desviación cuadrática media.

Se llama *desviación cuadrática media* de una magnitud aleatoria X la raíz cuadrada de la dispersión:

$$\sigma(X) = \sqrt{D(X)}.$$

Se demuestra fácilmente que la dispersión tiene una dimensión igual al cuadrado de la dimensión de la magnitud aleatoria. Puesto que la desviación cuadrática media es igual a la raíz cuadrada de la dispersión, la dimensión de $\sigma(X)$ coincide con la dimensión de X. Por eso, cuando se quiere que la estimación de la dispersión tenga la dimensión de la magnitud aleatoria, se calcula la desviación cuadrática media y no la dispersión.

Por ejemplo, si X se expresa en metros lineales, $\sigma(X)$ se expresará también en metros lineales y $D(X)$, en metros cuadrados.

Ejemplo. La magnitud aleatoria X está prefijada por la ley de distribución

X	2	3	10
p	0,1	0,4	0,5.

Hallar la desviación cuadrática media $\sigma(X)$.

SOLUCION. Hallamos la esperanza matemática de X:

$$M(X) = 2\cdot 0{,}1 + 3\cdot 0{,}4 + 10\cdot 0{,}5 = 6{,}4.$$

Hallamos la esperanza matemática de X^2:

$$M(X^2) = 2^2\cdot 0{,}1 + 3^2\cdot 0{,}4 + 10^2\cdot 0{,}5 = 54.$$

Hallamos la dispersión:

$$D(X) = M(X^2) - [M(X)]^2 = 54 - 6{,}4^2 = 13{,}04.$$

La desviación cuadrática media buscada es

$$\sigma(X) = \sqrt{D(X)} = \sqrt{13{,}04} \simeq 3{,}61.$$

§ 8. Desviación cuadrática media de la suma de magnitudes aleatorias mutuamente independientes

Supongamos que se conocen las desviaciones cuadráticas medias de varias magnitudes aleatorias mutuamente independientes. ¿Cómo hallar la desviación cuadrática media de la suma de estas magnitudes? El siguiente teorema da la respuesta a esta pregunta.

Teorema. *La desviación cuadrática media de la suma de un número finito de magnitudes aleatorias mutuamente independientes es igual a la raíz cuadrada de la suma de los cuadrados de las desviaciones cuadráticas medias de estas magnitudes:*

$$\sigma(X_1 + X_2 \ldots + X_n) = \\ = \sqrt{\sigma^2(X_1) + \sigma^2(X_2) + \ldots + \sigma^2(X_n)}.$$

DEMOSTRACION. Designamos por X la suma de las magnitudes mutuamente independientes consideradas:

$$X = X_1 + X_2 + \ldots + X_n.$$

Puesto que la dispersión de la suma de varias magnitudes aleatorias mutuamente independientes es igual a la suma de las dispersiones de los sumandos (§ 5, corolario 1), entonces

$$D(X) = D(X_1) + D(X_2) + \ldots + D(X_n).$$

De donde

$$\sqrt{D(X)} = \sqrt{D(X_1) + D(X_2) + \ldots + D(X_n)},$$

o bien, finalmente

$$\sigma(X) = \sqrt{\sigma^2(X_1) + \sigma^2(X_2) + \ldots + \sigma^2(X_n)}.$$

§ 9. Magnitudes aleatorias mutuamente independientes igualmente distribuidas

Ya se sabe que por la ley de distribución se pueden hallar las características numéricas de una magnitud aleatoria. De aquí se deduce que si varias magnitudes aleatorias tienen iguales distribuciones, sus características numéricas son idénticas.

Examinemos n magnitudes aleatorias mutuamente independientes $X_1, X_2, \ldots X_n$ que tienen iguales distribuciones

y, por lo tanto, también iguales características (esperanza matemática, dispersión, etc.). El estudio de las características numéricas de la media aritmética de estas magnitudes presenta mayor interés, de lo que nos ocuparemos en el presente párrafo.

La media aritmética de las magnitudes aleatorias examinadas la designamos por $\overline{X}$:

$$\overline{X}=\frac{X_1+X_2+\ldots+X_n}{n}.$$

Los tres postulados que siguen más adelante establecen la relación entre las características numéricas de la media aritmética $\overline{X}$ y las correspondientes características de cada magnitud por separado.

1. *La esperanza matemática de la media aritmética de magnitudes aleatorias mutuamente independientes igualmente distribuidas es igual a la esperanza matemática a de cada una de las magnitudes:*

$$M(\overline{X})=a.$$

DEMOSTPACIÓN. Utilizando las propiedades de la esperanza matemática (el factor constante se puede sacar fuera del signo de esperanza matemática; la esperanza matemática de una suma es igual a la suma de las esperanzas matemáticas de los sumandos), tendremos:

$$M(\overline{X})=M\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)=\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}.$$

Teniendo en cuenta que la esperanza matemática de cada una de las magnitudes, por los datos, es igual a a, obtenemos

$$M(\overline{X})=\frac{na}{n}=a.$$

2. *La dispersión de la media aritmética de n magnitudes aleatorias mutuamente independientes igualmente distribuidas es n veces menor que la dispersión D de cada una de las magnitudes:*

$$D(\overline{X})=\frac{D}{n}. \qquad (*)$$

DEMOSTRACIÓN. Utilizando las propiedades de la dispersión (el factor constante se puede sacar fuera del signo

de dispersión, elevándolo al cuadrado; la dispersión de la suma de magnitudes independientes es igual a la suma de las dispersiones de los sumandos), tenemos

$$D(\overline{X}) = D\left(\frac{X_1+X_2+\ldots+X_n}{n}\right) = \\ = \frac{D(X_1)+D(X_2)+\ldots+D(X_n)}{n^2}.$$

Tomando en consideración que la dispersión de cada una de las magnitudes es igual a D, obtenemos:

$$D(\overline{X}) = \frac{nD}{n^2} = \frac{D}{n}.$$

3. *La desviación media cuadrática de la media aritmética de n magnitudes aleatorias mutuamente independientes igualmente distribuidas es $\sqrt{n}$ veces menor que la desviación cuadrática media σ de cada magnitud:*

$$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}. \qquad (**)$$

DEMOSTRACION. Puesto que $D(X) = \frac{D}{n}$, la desviación cuadrática media de $\overline{X}$ es igual a

$$\sigma(\overline{X}) = \sqrt{D(\overline{X})} = \sqrt{\frac{D}{n}} = \frac{\sqrt{D}}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}.$$

De las fórmulas (*) y (**) se hace la siguiente deducción general: recordando que la dispersión y la desviación cuadrática media sirven como medidas de la dispersión de una magnitud aleatoria, inferimos que la media aritmética de un número suficientemente grande de magnitudes aleatorias mutuamente independientes tiene una dispersión bastante menor que cada magnitud individual.

Aclaremos con un ejemplo el valor de esta deducción para la práctica.

Ejemplo. Generalmente para medir cierta magnitud física se realizan varias mediciones y luego se halla la media aritmética de los números obtenidos que se toma como valor aproximado de la magnitud que se mide. Suponiendo que las mediciones se realizan en iguales condiciones, demostrar que:

a) la media aritmética da un resultado más fiable que las mediciones individuales;

b) al aumentar el número de mediciones la fiabilidad de este resultado crece.

SOLUCION. a) Se sabe que las mediciones individuales no dan idénticos valores de la magnitud a medir. El resultado de cada medición depende de muchas causas fortuitas (variación de la temperatura, oscilaciones del instrumento, etc.) que no pueden ser tomadas en consideración con anticipación.

Por eso, tenemos derecho de considerar los resultados posibles de n mediciones individuales como magnitudes aleatorias X_1, X_2, . . ., X_n (el subíndice indica el número de la medición). Estas magnitudes tienen igual distribución de las probabilidades (las mediciones se efectúan con igual metodología y los mismos instrumentos) y, por lo tanto, idénticas características numéricas; además, ellas son mutuamente independientes (el resultado de cada medición individual no depende de las restantes mediciones).

Ya sabemos que la media aritmética de tales magnitudes tiene menos dispersión que cada magnitud por separado. En otras palabras, la media aritmética resulta más próxima al valor real de la magnitud a medir que el resultado de una medición individual. Esto denota precisamente que, la media aritmética de varias mediciones da un resultado más fiable que la medición individual.

b) Se sabe que, al aumentar el número de magnitudes aleatorias individuales, la dispersión de la media aritmética disminuye. Esto significa que, al aumentar el número de mediciones, la media aritmética de varias mediciones se diferencia mucho menos del valor real de la magnitud que se mide. En consecuencia, incrementando el número de mediciones se obtiene un resultado más fiable.

Por ejemplo, si la desviación cuadrática media de una medición individual es $\sigma = 6$ m y en total se han realizado $n = 36$ mediciones, la desviación cuadrática media de la media aritmética de estas mediciones es igual solamente a 1 m. En efecto,

$$\sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = 1.$$

Vemos que la media aritmética de varias mediciones, como era de esperar, resultó más próxima al valor real de la magnitud a medir que el resultado de una medición individual.

§ 10. Noción de momentos de distribución

Consideremos la magnitud aleatoria discreta X, prefijada por la ley de distribución

X	1	2	5	100
p	0,6	0,2	0,19	0,01.

Hallamos la esperanza matemática de X:

$$M(X) = 1\cdot 0{,}6 + 2\cdot 0{,}2 + 5\cdot 0{,}19 + 100\cdot 0{,}01 = 2{,}95.$$

Escribimos la ley de distribución de X^2:

X^2	1	4	25	10 000
p	0,06	0,02	0,19	0,01.

Hallamos la esperanza matemática de X^2:

$$M(X^2) = 1\cdot 0{,}6 + 4\cdot 0{,}2 + 25\cdot 0{,}19 + 10\,000\cdot 0{,}01 = 106{,}15.$$

Como vemos, $M(X^2)$ es bastante mayor que $M(X)$. Esto se debe a que, después de elevar al cuadrado el valor posible de la magnitud X^2 que corresponde al valor $x = 100$ de la magnitud X, se ha hecho igual a 10 000, es decir, aumentó considerablemente; la probabilidad de esta magnitud es pequeña (0,01).

De este modo, el paso de $M(X)$ a $M(X^2)$ permitió considerar mejor la influencia en la esperanza matemática de aquel valor posible que es grande y tiene pequeña probabilidad. Está claro que si la magnitud de X tuviese unos cuantos valores grandes y poco probables, el paso a la magnitud X^2, y más aún a las magnitudes X^3, X^4, etc., permitiría «ampliar» aún más el «papel» de estos valores posibles grandes, pero poco probables. He aquí porque resulta conveniente considerar la esperanza matemática de potencia positiva entera de una magnitud aleatoria (no solamente discreta, sino también continua).

La esperanza matemática de la magnitud X^k se llama *momento inicial de orden k* de una magnitud aleatoria X:

$$\nu_k = M(X^k).$$

En particular,

$$\nu_1 = M(X),$$
$$\nu_2 = M(X^2).$$

Utilizando estos momentos, la fórmula para el cálculo de la dispersión $D(X) = M(X^2) - [M(X)]^2$ se puede escribir así:

$$D(X) = \nu_2 - \nu_1^2. \qquad (*)$$

Además de los momentos de la magnitud aleatoria X, conviene examinar los momentos de la desviación $X - M(X)$.

La esperanza matemática de la magnitud $(X - M(X))^k$ *se llama momento central de orden k de una magnitud aleatoria X*:

$$\mu_k = M[(X - M(X))^k].$$

En particular,

$$\mu_1 = M[(X - M(X))] = 0, \qquad (**)$$

$$\mu_2 = M[(X - M(X))^2] = D(X). \qquad (***)$$

Fácilmente se deducen las correlaciones que vinculan los momentos inicial y central.

Por ejemplo, comparando (*) y (***), obtenemos

$$\mu_2 = \nu_2 - \nu_1^2.$$

Partiendo de la definición de momento central y utilizando las propiedades de la esperanza matemática se obtienen fácilmente las fórmulas:

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3,$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.$$

Raramente se utilizan los momentos de órdenes más altos.

Nota. Los momentos estudiados aquí se llaman *teóricos*. A diferencia de los momentos teóricos, los momentos que se calculan por los datos de observaciones, se llaman *empíricos*. Las definiciones de los momentos empíricos están dadas más adelante (cáp. XVII, § 2).

Problemas

1. Conocidas las dispersiones de dos magnitudes aleatorias independientes: $D(X) = 4$ y $D(Y) = 3$. Hallar la dispersión de la suma de estas magnitudes.

Respuesta 7.

2. La dispersión de una magnitud aleatoria X es igual a 5. Hallar la dispersión de las siguientes magnitudes: a) $X - 1$; b) $-2X$; c) $3X + 6$.

Respuesta a) 5; b) 20; c) 45.

3. La magnitud aleatoria X toma solamente dos valores $+C$ y $-C$, cada uno con probabilidad 0,5. Hallar la dispersión de esta magnitud.

Respuesta C^2.

4. Hallar la dispersión de una magnitud aleatoria conociendo su ley de distribución

X	0,1	2	10	20
p	0,4	0,2	0,15	0,25.

Respuesta 67,0404.

5. La magnitud aleatoria X puede tomar dos valores posibles: x_1 con probabilidad 0,3 y x_2 con probabilidad 0,7; además $x_2 > x_1$. Hallar x_1 y x_2 sabiendo que $M(X) = 2,7$ y $D(X) = 0,21$.

Respuesta $x_1 = 2$; $x_2 = 3$.

6. Hallar la dispersión de la magnitud aleatoria X, es decir, del número de apariciones del suceso A en dos pruebas independientes, si $M(X) = 0,8$.

Indicación. Escribir la ley binomial de distribución de las probabilidades del número de apariciones del suceso A en dos tentativas independientes.

Respuesta 0,48.

7. Se ensaya un dispositivo compuesto de cuatro aparatos que trabajan independientemente. Las probabilidades de que fallen los aparatos son: $p_1 = 0,3$; $p_2 = 0,4$; $p_3 = 0,5$; $p_4 = 0,6$. Hallar la esperanza matemática y la dispersión del número de aparatos que fallan.

Respuesta 1,8; 0,94.

8. Hallar la dispersión de la magnitud aleatoria X, es decir, del número de apariciones de un suceso en 100 pruebas independientes, en cada una de las cuales la probabilidad de que ocurra el suceso es igual a 0,7.

Respuesta 21.

9. La dispersión de una magnitud aleatoria es $D(X) = 6,25$. Hallar la desviación cuadrática media $\sigma(X)$.

Respuesta 2,5.

10. Una magnitud aleatoria está prefijada por la ley de distribución

X	2	4	8
p	0,1	0,5	0,4.

Hallar la desviación cuadrática media de esta magnitud.

Respuesta 2,2.

11. La dispersión de cada una de las 9 magnitudes aleatorias mutuamente independientes igualmente distribuidas, es igual a 36. Hallar la dispersión cuadrática media de estas magnitudes.

Respuesta 4.

12. La desviación cuadrática media de cada una de las 16 magnitudes aleatorias mutuamente independientes igualmente distribuidas es igual a 10. Hallar la desviación cuadrática media de la media aritmética de estas magnitudes.

Respuesta 2,5.

Capítulo noveno

LEY DE LOS GRANDES NUMEROS

§ 1. Observaciones preliminares

Como se sabe, no se puede predecir con certeza, cuál de los valores posibles toma la magnitud aleatoria al final de la prueba; esto depende de muchas causas fortuitas, las que no estamos en condición de considerar. Aparentemente, dado que disponemos de datos muy modestos sobre cada magnitud aleatoria, difícilmente se puede establecer la ley o regularidad de comportamiento y la suma de un número suficientemente grande de magnitudes aleatorias. En realidad, esto no es así. Resulta que para ciertas condiciones relativamente amplias el comportamiento total de un número bastante grande de magnitudes aleatorias casi pierde el carácter aleatorio y deviene regular.

Para la práctica es muy importante saber las condiciones, al cumplirse las cuales, la acción global de muchísimas causas fortuitas da lugar a un resultado casi independiente del caso, ya que permite pronosticar la marcha de los fenómenos. Precisamente estas condiciones se indican en los teoremas que llevan el nombre general de ley de los grandes números. Entre ellos se encuentran los teoremas de Chebishev y de Bernoulli (existen otros teoremas que aquí no se consideran). El teorema de Chebishev es la ley de los grandes números más general, en tanto que el teorema de Bernoulli, es la elemental.

Para demostrar estos teoremas nos servimos de la desigualdad de Chebishev.

§ 2. Desigualdad de Chebishev

La desigualdad de Chebishev se cumple para las magnitudes aleatorias discretas y continuas. A fin de simplificar nos limitaremos a la demostración de esta desigualdad para las magnitudes discretas.

Examinemos la magnitud aleatoria discreta X prefijada por la tabla de distribución:

$$\begin{array}{cccc} X & x_1 & x_2 \ldots & x_n \\ p & p_1 & p_2 \ldots & p_n. \end{array}$$

Vamos a estimar la probabilidad de que la desviación de una magnitud aleatoria respecto de su esperanza matemática no supera en valor absoluto el número positivo ε. Si ε es bastante pequeño, estimaremos, en consecuencia, la probabilidad de que X toma valores suficientemente próximos a su esperanza matemática. P. L. Chebishev demostró la desigualdad, que permite dar la estimación que nos interesa.

DESIGUALDAD DE CHEBISHEV. *La probabilidad de que la desviación de la magnitud aleatoria X respecto de su esperanza matemática es menor en valor absoluto que un número positivo ε no menor que $1 - \frac{D(X)}{\varepsilon^2}$:*

$$P(|X - M(X)| < \varepsilon) \geqslant 1 - \frac{D(X)}{\varepsilon^2}.$$

DEMOSTRACIÓN. Puesto que los sucesos compuestos por el cumplimiento de las desigualdades $|X - M(X)| < \varepsilon$ y $|X - M(X)| \geqslant \varepsilon$ son opuestos, la suma de sus probabilidades es igual a la unidad, es decir,

$$P(|X - M(X)| < \varepsilon) + P(|X - M(X)| \geqslant \varepsilon) = 1.$$

De aquí, la probabilidad que nos interesa es

$$P(|X - M(X)| < \varepsilon) = 1 - P(|X - M(X)| \geqslant \varepsilon). \qquad (*)$$

Vemos que el problema se reduce al cálculo de la probabilidad $P(|X - M(X)| \geqslant \varepsilon)$.

Escribimos la expresión de la dispersión de la magnitud aleatoria X:

$$D(X) = [x_1 - M(X)]^2 p_1 + [x_2 - M(X)]^2 p_2 + \ldots$$
$$\ldots + [x_n - M(X)]^2 p_n.$$

Evidentemente, todos los sumandos de esta suma no son negativos.

Omitimos los sumandos, en los cuales $|x_j - M(X)| < \varepsilon$ (para los sumandos que quedan $|x_j - M(X)| \geqslant \varepsilon$), debido a lo cual la suma sólo puede reducirse. Vamos a considerar para certeza que se han suprimido los k primeros sumandos (sin alterar la comunidad, se puede suponer que en la tabla de distribución los valores posibles están enumerados, precisamente, en este orden). Por lo tanto,

$$D(X) \geqslant [x_{k+1} - M(X)]^2 p_{k+1} + [x_{k+2} - \\ - M(X)]^2 p_{k+2} + \ldots + [x_n - M(X)]^2 p_n.$$

Conviene hacer notar que ambos miembros de la desigualdad $|x_j - M(X)| \geqslant \varepsilon$ $(j = k+1, k+2, \ldots, n)$ son positivos, por eso elevándolos al cuadrado obtenemos la desigualdad equivalente $|x_j - M(X)|^2 \geqslant \varepsilon^2$. Utilizamos esta observación y, sustituyendo en la suma cada uno de los factores $|x_j - M(X)|^2$ por el número ε^2 (en este caso la desigualdad sólo puede acrecentarse), obtenemos

$$D(X) \geqslant \varepsilon^2 (p_{k+1} + p_{k+2} + \ldots + p_n). \qquad (**)$$

Por el teorema de la adición la suma de las probabilidades $p_{k+1} + p_{k+2} + \ldots + p_n$ es la probabilidad de que X tome uno, indiferentemente cual, de los valores $x_{k+1}, x_{k+2}, \ldots, x_n$, y para cualesquiera de ellos la desviación satisfará la desigualdad $|x_j - M(X)| \geqslant \varepsilon$. De aquí se deduce que la suma $p_{k+1} + p_{k+2} + \ldots + p_n$ expresa la probabilidad

$$P(|X - M(X)| \geqslant \varepsilon).$$

Esta consideración permite escribir la desigualdad (**) así:

$$D(X) \geqslant \varepsilon^2 \cdot P(|X - M(X)| \geqslant \varepsilon),$$

o bien

$$P(|X - M(x)| \geqslant \varepsilon) \leqslant \frac{D(X)}{\varepsilon^2}. \qquad (***)$$

Poniendo la (***) en (*), finalmente obtenemos

$$P(|X - M(X)| < \varepsilon) \geqslant 1 - \frac{D(X)}{\varepsilon^2}$$

lo que se quería demostrar.

Nota. Para la práctica la desigualdad de Chebishev tiene un valor limitado, puesto que de ordinario da una estimación aproximada y a veces trivial (que no presenta interés). Por ejemplo, si $D(X) > \varepsilon^2$ y,

por lo tanto $\frac{D(X)}{\varepsilon^2} > 1$, tendremos que $1 - \frac{D(X)}{\varepsilon^2} < 0$; por consiguiente, en este caso la desigualdad de Chebishev indica sólo que la probabilidad de la desviación no es negativa, pero esto desde ya es evidente, ya que cualquier probabilidad se expresa por un número no negativo.

El valor teórico de la desigualdad de Chebishev es muy grande. A continuación utilizamos esta desigualdad para deducir el teorema de Chebishev.

§ 3. Teorema de Chebishev

Teorema de Chebishev. *Si $X_1, X_2, \ldots, X_n$ son magnitudes aleatorias independientes de dos en dos, además sus dispersiones están uniformemente limitadas (no superan el número constante C), por más pequeño que sea el número positivo ε, la probabilidad de la desigualdad*

$$\left| \frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{M(X_1) + M(X_2) + \ldots + M(X_n)}{n} \right| < \varepsilon$$

será tan próxima a la unidad como se quiera, si el número de magnitudes aleatorias es bastante grande.

En otras palabras, de los datos del teorema

$$\lim_{n \to \infty} P\left(\left| \frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{M(X_1) + M(X_2) + \ldots + M(X_n)}{n} \right| < \varepsilon \right) = 1.$$

Por consiguiente, el teorema de Chebishev establece que si se examina un número bastante grande de magnitudes aleatorias independientes que tienen dispersiones limitadas, casi con certeza se puede considerar el suceso consistente en que la desviación media aritmética de las magnitudes aleatorias respecto de la media aritmética de sus esperanzas matemáticas será en valor absoluto tan pequeño como se quiera.

DEMOSTRACIÓN. Introducimos en el examen una nueva magnitud aleatoria, o sea, la media aritmética de las magnitudes aleatorias

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Hallamos la esperanza matemática de $\overline{X}$. Utilizando las propiedades de la esperanza matemática (un factor constante

se puede sacar fuera del signo de esperanza matemática, la esperanza matemática de la suma es igual a la suma de las esperanzas matemáticas de los sumandos), obtenemos

$$M\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)=\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}. \qquad (*)$$

Aplicando a la magnitud $\overline{X}$ la desigualdad de Chebishev, tenemos

$$P\left[\left|\frac{X_1+X_2+\ldots+X_n}{n}-M\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)\right|<\varepsilon\right]\geqslant$$

$$\geqslant 1-\frac{D\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)}{\varepsilon^2};$$

o bien, teniendo en cuenta la correlación (*),

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\right.\right.$$

$$\left.\left.-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|<\varepsilon\right)\geqslant 1-$$

$$-\frac{D\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)}{\varepsilon^2}. \qquad (**)$$

Utilizando las propiedades de la dispersión (un factor constante se puede sacar fuera del signo de dispersión, elevándolo al cuadrado; la dispersión de la suma de magnitudes aleatorias independientes es igual a la suma de las dispersiones de los sumandos), obtenemos

$$D\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)=\frac{D(X_1)+D(X_2)+\ldots+D(X_n)}{n^2}.$$

Según la condición las dispersiones de todas las magnitudes aleatorias están limitadas por el número constante C, es decir, se cumplen las desigualdades:

$$D(X_1)\leqslant C;\quad D(X_2)\leqslant C;\ \ldots;\ D(X_n)\leqslant C,$$

por eso,

$$\frac{D(X_1)+D(X_2)+\ldots+D(X_n)}{n^2}\leqslant\frac{C+C+\ldots+C}{n^2}=\frac{nC}{n^2}=\frac{C}{n}$$

De este modo,

$$D\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)\leqslant\frac{C}{n}. \qquad (***)$$

Poniendo el segundo miembro de (***) en la desigualdad (**) (por lo cual esta última sólo puede ser acrecentada), tenemos

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\right.\right.$$
$$\left.\left.-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|<\varepsilon\right)\geqslant 1-\frac{C}{n\varepsilon^2}.$$

De aquí, pasando al límite para $n\to\infty$, obtenemos

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\right.\right.$$
$$\left.\left.-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|<\varepsilon\right)\geqslant 1.$$

Por último, teniendo en cuenta que la probabilidad no puede ser mayor que la unidad, finalmente podemos escribir

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-\right.\right.$$
$$\left.\left.-\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}\right|<\varepsilon\right)=1.$$

Por lo que el teorema queda demostrado.

Al formular el teorema de Chebishev hemos supuesto que las magnitudes aleatorias tienen diferentes esperanzas matemáticas. En la práctica generalmente ocurre que las magnitudes aleatorias tienen igual esperanza matemática. Es evidente que si se admite nuevamente que las dispersiones de estas magnitudes están limitadas, es posible aplicar a ellas el teorema de Chebishev.

Designemos por a la esperanza matemática de cada una de las magnitudes aleatorias; en el caso a examinar la media aritmética de las esperanzas matemáticas, como se aprecia fácilmente, también es igual a a.

Podemos formular el teorema de Chebishev para el caso particular estudiado.

Si X_1, X_2, ..., X_n son magnitudes aleatorias independientes de dos en dos, con igual esperanza matemática a y si la dispersión de estas magnitudes está uniformemente limitada, por más pequeño que sea el número $\varepsilon>0$, la probabilidad de la desigualdad

$$\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|<\varepsilon$$

será tan próxima a la unidad como se quiera, si el número de magnitudes aleatorias es suficientemente grande.

En otras palabras, por las condiciones del teorema tendrá lugar la igualdad

$$\lim_{n \to \infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|<\varepsilon\right)=1.$$

§ 4. Esencia del teorema de Chebishev

La esencia del teorema demostrado es: aunque las magnitudes aleatorias independientes individuales pueden tomar valores lejanos de sus esperanzas matemáticas, la media aritmética de un número bastante grande de magnitudes aleatorias con gran probabilidad toma valores próximos a un número constante determinado, precisamente al número $\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n}$ (o al número a en el caso particular). En otras palabras, algunas magnitudes aleatorias pueden tener una dispersión considerable y su media aritmética es de poca dispersión.

De este modo, no se puede pronosticar con certeza, qué valor posible toma cada una de las magnitudes aleatorias, pero sí se puede predecir qué valor toma su media aritmética.

Por consiguiente, la media aritmética de un número suficientemente grande de magnitudes aleatorias independientes (cuyas dispersiones están uniformemente limitadas) *pierde el carácter de magnitud aleatoria*. Esto se debe a que las desviaciones de cada una de las magnitudes de sus esperanzas matemáticas pueden ser tanto positivas, como negativas, y en la media aritmética se excluyen mutuamente.

El teorema de Chebishev se cumple no sólo para las magnitudes aleatorias discretas, sino también para las continuas; éste es un ejemplo preciso que confirma la validez de la doctrina del materialismo dialéctico sobre la relación entre la casualidad y la necesidad.

§ 5. Valor práctico del teorema de Chebishev

Expondremos algunos ejemplos de aplicación del teorema de Chebishev a la resolución de problemas prácticos.

Generalmente para medir cierta magnitud física se realizan varias mediciones y su media aritmética se toma como

la dimensión buscada. ¿En qué condiciones este método de medición puede considerarse correcto? La respuesta la da el teorema de Chebishev (su caso particular).

En efecto, examinemos los resultados de cada medición como magnitudes aleatorias $X_1, X_2, \ldots, X_n$. A estas magnitudes se les puede aplicar el teorema de Chebishev si: 1) son independientes de dos en dos, 2) tienen igual esperanza matemática, 3) sus dispersiones están uniformemente limitadas.

La primera exigencia se cumple si el resultado de cada medición no depende de los resultados de las demás.

La segunda exigencia se cumple si las mediciones se han realizado sin errores sistemáticos (de un signo). En este caso, las esperanzas matemáticas de todas las magnitudes aleatorias son idénticas e iguales a la dimensión verdadera a.

La tercera exigencia se cumple si el instrumento garantiza una determinada precisión de las mediciones. Aunque en este caso los resultados de las mediciones individuales son distintos, su dispersión está limitada.

Si se cumplen todos los requisitos indicados, tenemos el derecho de aplicar el teorema de Chebishev a los resultados de las mediciones: para n suficientemente grande la probabilidad de la desigualdad

$$\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|<\varepsilon$$

se aproxima a la unidad tanto como se quiera. En otras palabras, para un número bastante grande de mediciones casi con certeza su media aritmética se diferencia tan poco como se quiera del valor verdadero de la magnitud que se mide.

Por consiguiente, el teorema de Chebishev indica las condiciones, para las cuales puede ser aplicado el método de medición descripto.

Sin embargo, es erróneo pensar que aumentando el número de mediciones se puede lograr una precisión tan grande como se quiera. El problema está en que el propio instrumento da indicaciones (lecturas) sólo con la precisión de $\pm\alpha$; por eso, cada uno de los resultados de las mediciones y, por lo tanto, también su media aritmética, se obtendrán con una exactitud no mayor que la precisión del instrumento.

El método muestral, profusamente aplicado en estadística y basado en el teorema de Chebishev, en esencia tiene como objeto juzgar por un muestreo aleatorio relativamente

pequeño, de todo el conjunto (conjunto general) de objetos investigados. Por ejemplo, mediante un pequeño puñado de fibras escogidas al azar de distintas partes de un fardo se determina la calidad de algodón de todo el fardo. Aunque el número de fibras en el puñado es considerablemente menor que en el fardo, el propio puñado contiene una cantidad suficientemente grande de fibras, calculada en cientos.

Un otro ejemplo puede ser la determinación de la calidad del grano por una pequeña muestra suya. También en este caso el número de granos escogidos al azar es pequeño en comparación con toda la masa de granos, pero de por sí es bastante grande.

De los ejemplos expuestos se puede deducir que el teorema de Chebishev tiene un valor inestimable para la práctica

§ 6. Teorema de Bernoulli

Supongamos que se realizan n experimentos independientes, en cada uno de los cuales la probabilidad de que ocurra el suceso A es igual a p. ¿Puede predecirse cuál será aproximadamente la frecuencia relativa de apariciones del suceso? El teorema demostrado por Jacobo Bernoulli (publicado en 1713) que se denominara «ley de los grandes números» y diera origen a la teoría de las probabilidades como ciencia, da una respuesta positiva a esta pregunta. La demostración de Bernoulli era compleja; en 1864, P. V. Chebishev dio una demostración sencilla.

Teorema de Bernoulli. *Si en cada una de las n pruebas independientes la probabilidad p de que ocurra un suceso A es constante, la probabilidad tan próxima a la unidad como se quiera, de que la desviación de la frecuencia relativa respecto de la probabilidad p será en valor absoluto tan pequeño como se quiera, si el número de pruebas es suficientemente grande.*

En otras palabras, si ε es un número tan pequeño como se quiera, al cumplirse las condiciones del teorema se obtendrá la igualdad

$$\lim_{n \to \infty} P\left(\left| \frac{m}{n} - p \right| < \varepsilon \right) = 1.$$

DEMOSTRACION. Designemos por X_1 una magnitud aleatoria discreta, es decir, el número de apariciones de un suceso en la primera prueba; por X_2, en la segunda, . . ., por X_n en la n-sima prueba.

Es evidente que cada una de las magnitudes puede tomar sólo dos valores: 1 (el suceso A ocurrió) con probabilidad p y 0 (el suceso no ocurrió) con probabilidad $1 - p = q$.

¿Se podrá aplicar el teorema de Chebishev a las magnitudes a estudiar? Se puede, si las magnitudes aleatorias son independientes de dos en dos y su dispersión está limitada. Ambas condiciones se cumplen. En efecto, la independencia de dos en dos de las magnitudes $X_1, X_2, \ldots, X_n$ se deduce de la independencia de las pruebas. La dispersión de cualquier magnitud X_i ($i = 1, 2, \ldots, n$) es igual al producto pq*, ya que $p + q = 1$, el producto pq no es mayor de $\frac{1}{4}$** y, por lo tanto, las dispersiones de todas las magnitudes están limitadas, por ejemplo, por el número $C = \frac{1}{4}$.

Aplicando el teorema de Chebishev (caso particular) a las magnitudes examinadas, tenemos

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n} - a\right| < \varepsilon\right) = 1.$$

Tomando en consideración que la esperanza matemática a de cada una de las magnitudes X_i (es decir, la esperanza matemática del número de apariciones del suceso en una prueba) es igual a la probabilidad p de que ocurra el suceso (ejemplo 2, pág. 80), obtenemos

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n} - p\right| < \varepsilon\right) = 1.$$

Queda por demostrar que $\frac{X_1+X_2+\ldots+X_n}{n}$ es igual a la frecuencia relativa $\frac{m}{n}$ de apariciones del suceso A en n pruebas. En realidad, cada una de las magnitudes $X_1, X_2, \ldots, X_n$ toma un valor igual a la unidad al ocurrir el suceso en la respectiva prueba; por consiguiente, la suma $X_1 + X_2 + \ldots + X_n$ es igual al número m de apariciones del suceso en n pruebas, o sea,

$$\frac{X_1+X_2+\ldots+X_n}{n} = \frac{m}{n}.$$

---

* Esto se deduce del § 6, cap. VIII, si se admite que $n = 1$.

** Se sabe que el producto de dos factores, cuya suma es una magnitud constante, tiene un valor máximo cuando los factores son iguales. Aquí la suma $p_i + q_i = 1$, es decir, constante, por eso, cuando $p_i = q_i = \frac{1}{2}$, el producto $p_i q_i$ tiene un valor máximo y es igual a $\frac{1}{4}$.

Teniendo en cuenta esta igualdad, finalmente obtenemos

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right) = 1.$$

Nota. Sería erróneo deducir a base del teorema de Bernoulli que aumentando el número de experimentos la frecuencia relativa tiende indefectiblemente a la probabilidad p; en otras palabras, del teorema de Bernoulli no resulta la igualdad $\lim\limits_{n\to\infty} \frac{m}{n}=p$. En el teorema se trata solamente sobre la *probabilidad* de que para un número bastante grande de experimentos la frecuencia relativa se diferenciará tan poco como se quiera de la probabilidad constante de aparición del suceso en cada prueba.

De ese modo, la convergencia de la frecuencia relativa $\frac{m}{n}$ a la probabilidad p se diferencia de la convergencia en el sentido del análisis ordinario. Para destacar esta diferencia se introduce el concepto de «convergencia de probabilidad». Más exactamente, la diferencia entre los tipos de convergencia señalados consiste en lo siguiente: si $\frac{m}{n}$ tiende a p, para $n\to\infty$, como límite en el sentido del análisis ordinario, a partir de cierto $n=N$ y para todos los valores siguientes de n, se cumple consecuentemente la desigualdad $\left|\frac{m}{n}-p\right|<\varepsilon$; si $\frac{m}{n}$ tiende según probabilidad a p cuando $n\to\infty$, para valores individuales de n la desigualdad puede dejar de ser válida.

Así, el teorema de Bernoulli, afirma que para $n\to\infty$ la frecuencia relativa tiende según *probabilidad* a p. Sucintamente el teorema de Bernoulli se escribe así:

$$\frac{m}{n} \xrightarrow[n\to\infty]{\text{probab}} p.$$

Como podemos apreciar, el teorema de Bernoulli aclara porqué la frecuencia relativa para un número bastante grande de experimentos tiene la propiedad de estabilidad y verifica la definición estadística de la probabilidad (cap I, §§ 5—6).

Problemas

1. Formular y escribir el teorema de Chebishev utilizando el concepto de «convergencia de probabilidad».

2. Utilizando la desigualdad de Chebishev estimar la probabilidad de que $|X-M(X)|<0{,}1$, si $D(X)=0{,}001$.

Respuesta $P\geqslant 0{,}9$.

3. Utilizando la desigualdad de Chebishev hallar ε. Dados $P(|X-M(X)|<\varepsilon)\geqslant 0{,}9$; $D(X)=0{,}004$.

Respuesta 0,2.

Capítulo décimo

FUNCION INTEGRAL DE DISTRIBUCION DE LAS PROBABILIDADES DE UNA MAGNITUD ALEATORIA

§ 1. Definición de la función integral de distribución*

Recordemos que una magnitud aleatoria discreta se prefija por la enumeración de todos sus valores posibles y sus probabilidades. Este método no es general: por ejemplo, no es aplicable para las magnitudes aleatorias continuas.

En efecto, examinemos una magnitud aleatoria X, cuyos valores posibles llenan ininterrumpidamente el intervalo (a, b). ¿Se puede componer la lista de todos los valores posibles de X? Evidentemente, esto no se puede realizar. Este ejemplo indica la conveniencia de dar un método general de prefijar todos los tipos de magnitudes aleatorias. Precisamente con este propósito se introduce la función integral de distribución.

| Supongamos que x es un número real. La probabilidad de un suceso consistente en que X toma un valor menor que x, es decir, la probabilidad del suceso $X < x$ la designamos por $F(x)$. Está claro que si x varía, en general, variará también $F(x)$, o sea, $F(x)$ es una función de x.

Se llama *función integral de distribución* la función $F(x)$ que determina para cada valor de x la probabilidad de que la magnitud aleatoria X tome un valor menor que x, es decir

$$F(x) = P(X < x).$$

Geométricamente esta igualdad se puede interpretar así: $F(x)$ es la probabilidad de que una magnitud aleatoria tome el valor representado sobre el eje numérico por un punto ubicado a la izquierda del punto x.

Ahora podemos dar una definición más precisa de una magnitud aleatoria continua: una magnitud aleatoria se llama continua, si su función integral de distribución $F(x)$ es continuamente diferenciable.

* Frecuentemente en lugar de «función integral» se utiliza el término «función de distribución».

§ 2. Propiedades de una función integral

Propiedad 1. *Los valores de una función integral corresponden a un segmento* [0; 1]:

$$0 \leqslant F(x) \leqslant 1.$$

DEMOSTRACION. Esta propiedad resulta de la definición de función integral como probabilidad: la probabilidad siempre es un número no negativo no mayor que la unidad.

Propiedad 2. *$F(x)$ es una función no decreciente, es decir,*

$$F(x_2) \geqslant F(x_1), \text{ si } x_2 > x_1.$$

DEMOSTRACION. Supongamos que $x_2 > x_1$. El suceso consistente en que X toma un valor menor que x_2, se puede dividir en los dos sucesos mutuamente excluyentes siguientes: 1) X toma un valor menor que x_1 con probabilidad $P(X < x_1)$; 2) X toma un valor que satisface la desigualdad $x_1 \leqslant X < x_2$, con probabilidad $P(x_1 \leqslant X < x_2)$. Por el teorema de la adición tenemos que

$$P(X < x_2) = P(X < x_1) + P(x_1 \leqslant X < x_2).$$

De donde

$$P(X < x_2) - P(X < x_1) = P(x_1 \leqslant X < x_2),$$

o bien

$$F(x_2) - F(x_1) = P(x_1 \leqslant X < x_2). \qquad (*)$$

Ya que cualquier probabilidad es un número no negativo, tendremos que $F(x_2) - F(x_1) \geqslant 0$, o bien $F(x_2) \geqslant F(x_1)$, lo que se quería demostrar.

Corolario 1. *La probabilidad de que una magnitud aleatoria tome un valor acotado en el intervalo (a, b), es igual al incremento de la función integral en este intervalo:*

$$P(a \leqslant X < b) = F(b) - F(a). \qquad (**)$$

Este importante corolario resulta de la fórmula (*), si se pone $x_2 = b$ y $x_1 = a$.

Ejemplo. Una magnitud aleatoria X está prefijada por la función integral:

$$F(x) = \begin{cases} 0 & \text{para } x \leqslant -1; \\ \frac{1}{4}x + \frac{1}{4} & \text{para } -1 < x \leqslant 3; \\ 1 & \text{para } x > 3. \end{cases}$$

Hallar la probabilidad de que como resultado de la prueba X toma un valor acotado en el intervalo (0; 2)

$$P(0 < X < 2) = F(2) - F(0).$$

Puesto que en el intervalo (0; 2) por los datos

$$F(x) = \frac{1}{4}x + \frac{1}{4},$$

entonces

$$F(2) - F(0) = \left[\frac{1}{4}\cdot 2 + \frac{1}{4}\right] - \left[\frac{1}{4}\cdot 0 + \frac{1}{4}\right] = \frac{1}{2}.$$

Por consiguiente,

$$P(0 < X < 2) = \frac{1}{2}.$$

Corolario 2. *La probabilidad de que una magnitud aleatoria continua X toma un valor determinado, es igual a cero.*

En efecto, poniendo en la fórmula (**), $a = x_1$, $b = x_1 + \Delta x$ tenemos

$$P(x_1 \leqslant X < x_1 + \Delta x) = F(x_1 + \Delta x) - F(x_1).$$

Hacemos tender Δx a cero. Dado que X es una magnitud aleatoria continua, la función $F(x)$ es continua. Debido a la continuidad de $F(x)$ en el punto x_1 la diferencia $F(x_1 + \Delta x) - F(x_1)$ también tenderá a cero, por lo tanto, $P(X = x_1) = 0$.

Utilizando esta tésis, se aprecia fácilmente la validez de las igualdades

$$P(a \leqslant X < b) = P(a < X < b) = \\ = P(a < X \leqslant b) = P(a \leqslant X \leqslant b). \qquad (***)$$

Por ejemplo, la igualdad $P(a < X \leqslant b) = P(a < X < b)$ se demuestra así:

$$P(a < X \leqslant b) = P(a < X < b) + \\ + P(X = b) = P(a < X < b).$$

De este modo, no presenta interés referirnos a la probabilidad de que una magnitud aleatoria continua tome un valor determinado, pero tiene sentido considerar la probabilidad de que esté acotado en el intervalo, digamos tan pequeño como se quiera.

Este hecho corresponde totalmente a las exigencias de los problemas prácticos. Por ejemplo, nos interesa la probabilidad de que las medidas de las piezas no exceden los límites admisibles, pero no se plantea la probabilidad de su coincidencia con la dimensión de proyecto.

Conviene hacer notar que sería incorrecto pensar que siendo la probabilidad $P(X = x_1)$ igual a cero significa que el suceso $X = x_1$ es imposible (claro está, si no nos limitamos a la definición clásica de la probabilidad). En efecto, como resultado de la prueba la magnitud aleatoria toma indefectiblemente uno de los valores posibles; en particular, este valor puede resultar igual a x_1.

Propiedad 3. *Si los valores posibles de una magnitud aleatoria están acotados en el intervalo (a, b), tendremos que*

$$1)\ F(x) = 0 \qquad \text{para } x \leqslant a;$$

$$2)\ F(x) = 1 \qquad \text{para } x \geqslant b.$$

DEMOSTRACION. 1) Supongamos que $x_1 \leqslant a$. En ese caso, el suceso $X < x_1$ es imposible (ya que X no toma valores menores que x_1 por dato) y, por lo tanto, su probabilidad es igual a cero.

2) Supongamos que $x_2 \geqslant b$. En ese caso el suceso $X < x_2$ es verdadero (ya que todos los valores posibles de X son menores que x_2) y, por consiguiente, su probabilidad es igual a la unidad.

Corolario. *Si los valores posibles de una magnitud aleatoria continua se encuentran en todo el eje x, son justas las siguientes correlaciones límites:*

$$\lim_{x \to -\infty} F(x) = 0; \qquad \lim_{x \to \infty} F(x) = 1.$$

§ 3. Gráfica de la función integral

Las propiedades demostradas permiten representar gráficamente la función integral de una magnitud aleatoria continua.

La gráfica está ubicada en una banda limitada por las rectas $y = 0$, $y = 1$ (primera propiedad).

Al crecer x en el intervalo (a, b), en el que están acotados todos los valores posibles de la magnitud aleatoria, la gráfica «sube» (segunda propiedad).

Cuando $x \leqslant a$ las ordenadas de la gráfica son iguales a cero; si $x \geqslant b$, las ordenadas de la gráfica son iguales a la unidad (tercera propiedad).

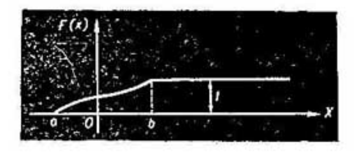


Fig. 1.

La gráfica de la función integral de una magnitud aleatoria continua está expuesta en la fig. 1.

Nota. Para una magnitud aleatoria discreta la gráfica de la función integral tiene forma escalonada. Verifiquemos esto con un ejemplo.

Ejemplo. Una magnitud aleatoria discreta X está prefijada por la siguiente tabla de distribución:

X	1	4	8
p	0,3	0,1	0,6.

Hallar la función integral y trazar su gráfica.

SOLUCION. 1°. Si $x \leqslant 1$, $F(x) = 0$ (tercera propiedad).

2°. Si $1 < x \leqslant 4$, $F(x) = 0{,}3$. En efecto, X puede tomar el valor 1 con la probabilidad 0,3.

3°. Si $4 < x \leqslant 8$, $F(x) = 0{,}4$. En realidad, si x_1 satisface la desigualdad $4 \leqslant x_1 \leqslant 8$, tendremos que $F(x_1)$ es igual a la probabilidad del suceso $X < x_1$, que puede ocurrir cuando X toma el valor 1 (la probabilidad de este suceso es igual a 0,3) o el valor 4 (la probabilidad de este suceso es igual a 0,1). Puesto que estos dos sucesos mutuamente excluyentes, por el teorema de la adición la probabilidad del suceso $X < x_1$ es igual a la suma de las probabilidades $0{,}3 + 0{,}1 = 0{,}4$.

4°. Si $x > 8$, $F(x) = 1$. En efecto, el suceso $X \leqslant 8$ es cierto y, por lo tanto, su probabilidad, igual a la unidad.

De este modo, analíticamente la función integral puede ser escrita así:

$$F(x)=\begin{cases}0 & \text{para} \quad x\leqslant 1,\\ 0{,}3 & \text{para} \quad 1<x\leqslant 4,\\ 0{,}4 & \text{para} \quad 4<x\leqslant 8,\\ 1 & \text{para} \quad x>8.\end{cases}$$

En la fig. 2 se muestra la gráfica de esta función.



Fig. 2.

Problemas

1. La magnitud aleatoria X está prefijada por la función integral:

$$F(x)=\begin{cases}0 & \text{para} \quad x\leqslant -1;\\ \frac{1}{3}x+\frac{1}{3} & \text{para} \quad -1<x\leqslant 2;\\ 1 & \text{para} \quad x>2.\end{cases}$$

Hallar la probabilidad de que como resultado de la prueba X toma un valor acotado en el intervalo (0; 1).

Respuesta $\frac{1}{3}$.

2. La magnitud aleatoria X está prefijada por la función integral:

$$F(x)=\begin{cases}0 & \text{para} \quad x\leqslant 2;\\ \frac{1}{2}x-1 & \text{para} \quad 2<x\leqslant 4;\\ 1 & \text{para} \quad x>4.\end{cases}$$

Hallar la probabilidad de que como resultado de la prueba X toma un valor acotado en el intervalo (2; 3).

Respuesta $\frac{1}{2}$.

3. La magnitud aleatoria discreta X está prefijada por la siguiente ley de distribución:

X	2	6	10
p	0,5	0,4	0,1.

Construir la gráfica de la función integral de esta magnitud.

Capítulo once

FUNCION DIFERENCIAL DE DISTRIBUCION DE LAS PROBABILIDADES DE UNA MAGNITUD ALEATORIA CONTINUA

§ 1. Definición de la función diferencial de distribución

Antes fijamos una magnitud aleatoria continua mediante la función integral. Ese método de prefijar no es el único. Una magnitud aleatoria continua se puede dar también utilizando la función diferencial de distribución de las probabilidades.

La derivada primera de la función integral se llama *función diferencial de distribución** $f(x)$:

$$f(x) = F'(x).$$

De la definición dada se deduce que la función integral es la primitiva para la función diferencial.

Cabe hacer notar que para describir la distribución de las probabilidades de una magnitud aleatoria discreta (discontinua) no es aplicable la función diferencial.

§ 2. Probabilidad de que una magnitud aleatoria continua caiga en un intervalo dado

Conociendo la función diferencial se puede calcular la probabilidad de que una magnitud aleatoria continua toma un valor correspondiente al intervalo dado. El cálculo se basa en el teorema siguiente.

Teorema. *La probabilidad de que la magnitud aleatoria continua X toma un valor perteneciente al intervalo* (a, b)

* Frecuentemente en vez de «función diferencial» se utiliza el término «densidad de la probabilidad».

es igual a una integral determinada de la función diferencial tomada entre los límites desde a hasta b:

$$P(a < X < b) = \int_a^b f(x)\,dx.$$

DEMOSTRACION. Utilicemos la correlación (**)(pág. 120):

$$P(a \leqslant X < b) = F(b) - F(a).$$

Por la fórmula de Newton—Leibniz

$$F(b) - F(a) = \int_a^b F'(x)\,dx = \int_a^b f(x)\,dx.$$

De este modo,

$$P(a \leqslant X < b) = \int_a^b f(x)\,dx.$$

Dado que $P(a \leqslant X < b) = P(a < X < b)$, finalmente obtenemos

$$P(a < X < b) = \int_a^b f(x)\,dx. \qquad (*)$$

El resultado obtenido geométricamente se puede interpretar así: la probabilidad de que una magnitud aleatoria

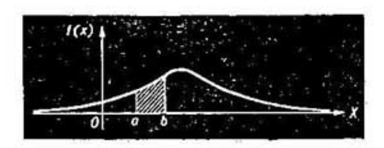


Fig. 3.

continua tome un valor perteneciente al intervalo (a, b), es igual a la superficie de un trapecio curvilíneo limitado por el eje x, la curva de distribución $f(x)$ y las rectas $x = a$ y $x = b$ (fig. 3).

Nota. En particular, si $f(x)$ es una función par y los extremos del intervalo son simétricos con respecto al origen de coordenadas, tendremos que

$$P(-a < X < a) = P(|X| < a) = 2\int_0^a f(x)\,dx.$$

Ejemplo. Dada la función diferencial de una magnitud aleatoria X:

$$f(x) = \begin{cases} 0 & \text{para} \quad x \leqslant 0; \\ 2x & \text{para} \quad 0 < x \leqslant 1; \\ 0 & \text{para} \quad x > 1. \end{cases}$$

Hallar la probabilidad de que debido a la prueba, X toma un valor perteneciente al intervalo (0,5; 1).

SOLUCION. La probabilidad buscada es

$$P(0{,}5 < X < 1) = 2\int_{0{,}5}^{1} x\,dx = x^2 \Big|_{0{,}5}^{1} = 1 - 0{,}25 = 0{,}75.$$

§ 3. Obtención de la función integral por la función diferencial conocida

Conociendo la función diferencial $f(x)$, se puede hallar la función integral $F(x)$ por la fórmula

$$F(x) = \int_{-\infty}^{x} f(x)\,dx.$$

En efecto, hemos designado por $F(x)$ la probabilidad de que la magnitud aleatoria tome un valor menor que x, es decir,

$$F(x) = P(X < x).$$

Evidentemente, la desigualdad $X < x$ se puede escribir en forma de una doble desigualdad $-\infty < X < x$, por lo tanto,

$$F(x) = P(-\infty < X < x). \qquad (*)$$

Suponiendo en la fórmula (*) (§ 2) $a = -\infty$, $b = x$, tenemos

$$P(-\infty < X < x) = \int_{-\infty}^{x} f(x)\,dx.$$

Por último, sustituyendo $P(-\infty < X < x)$ por $F(x)$, debido a (*), finalmente obtenemos

$$F(x)=\int_{-\infty}^{x} f(x)\,dx.$$

De este modo, conociendo la función diferencial de distribución, se puede hallar la función integral. Sin duda, cono-

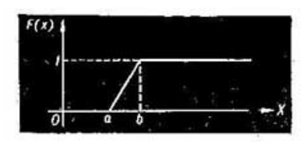


Fig. 4.

ciendo la función integral se puede hallar la función diferencial, es decir,

$$f(x) = F'(x).$$

Ejemplo. Dada la función diferencial, hallar la función integral

$$f(x)=\begin{cases} 0 & \text{para} \quad x\leqslant a, \\ \dfrac{1}{b-a} & \text{para} \quad a< x\leqslant b, \\ 0 & \text{para} \quad x>b. \end{cases}$$

Trazar la gráfica de la función hallada.

SOLUCION. Utilizamos la fórmula $F(x)=\int_{-\infty}^{x} f(x)\,dx$. Si $x\leqslant a$, tendremos que $f(x)=0$ y, por lo tanto, $F(x)=0$. Si $a<x\leqslant b$, $f(x)=\frac{1}{b-a}$ y, en consecuencia,

$$F(x)=\int_{-\infty}^{x} f(x)\,dx=\int_{-\infty}^{a} 0\,dx+\int_{a}^{x} \frac{1}{b-a}\,dx=\frac{x-a}{b-a}.$$

Si $x > b$, entonces

$$F(x) = \int_{-\infty}^{a} 0\,dx + \int_{a}^{b} \frac{dx}{b-a} + \int_{b}^{x} 0\,dx = \frac{b-a}{b-a} = 1.$$

De este modo, la función integral buscada puede ser escrita analíticamente así:

$$F(x) = \begin{cases} 0 & \text{para} \quad x \leqslant a, \\ \frac{x-a}{b-a} & \text{para} \quad a < x \leqslant b, \\ 1 & \text{para} \quad x > b. \end{cases}$$

La gráfica de esta función está representada en la fig. 4.

§ 4. Propiedades de la función diferencial

Propiedad 1. *La función diferencial no es negativa:*

$$f(x) \geqslant 0.$$

DEMOSTRACION. La función integral es una función no decreciente, por lo tanto, su derivada $F'(x) = f(x)$ es una función no negativa.

Geométricamente esta propiedad significa que los puntos pertenecientes a la gráfica de la función diferencial están ubicados o encima del eje x, o bien en ese eje.

Conviene hacer notar que la gráfica de la función diferencial se llama *curva de distribución*.

Propiedad 2. *La integral impropia de la función diferencial en los límites desde $-\infty$ hasta ∞ es igual a la unidad:*

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

DEMOSTRACION. La integral impropia $\int_{-\infty}^{\infty} f(x)\,dx$ expresa la probabilidad del suceso consistente en que la magnitud aleatoria toma un valor perteneciente al intervalo $(-\infty, \infty)$. Evidentemente, tal suceso es cierto y, por lo tanto, su probabilidad es igual a la unidad.

Geométricamente esto significa que la superficie del trapecio curvilíneo limitado por el eje x y la curva de distribución, es igual a la unidad.

En particular, si todos los valores posibles de la magnitud aleatoria pertenecen al intervalo (a, b), entonces

$$\int_a^b f(x)\,dx = 1.$$

Ejemplo. La función diferencial de distribución de la magnitud aleatoria X está prefijada por la igualdad

$$f(x) = \frac{a}{e^{-x}+e^{x}}.$$

Hallar el parámetro constante a.

SOLUCION. La función diferencial debe satisfacer la condición $\int_{-\infty}^{\infty} f(x)\,dx = 1$, por eso, se necesita que se cumpla la igualdad

$$a\int_{-\infty}^{\infty} \frac{dx}{e^{-x}+e^{x}} = 1.$$

De donde

$$a = \frac{1}{\int_{-\infty}^{\infty} \frac{dx}{e^{-x}+e^{x}}}.$$

Hallamos la integral indefinida

$$\int \frac{dx}{e^{-x}+e^{x}} = \int \frac{e^{x}\,dx}{1+e^{2x}} = \operatorname{arctg} e^{x}.$$

Calculamos la integral impropia

$$\int_{-\infty}^{\infty} \frac{dx}{e^{-x}+e^{x}} = \lim_{b\to\infty}\int_b^0 \frac{dx}{e^{-x}+e^{x}} + \lim_{c\to\infty}\int_0^c \frac{dx}{e^{-x}+e^{x}} =$$

$$= \lim_{b\to-\infty}(-\operatorname{arctg} e^{b}) + \lim_{c\to\infty}(\operatorname{arctg} e^{c}) = \frac{\pi}{2}.$$

De este manera, el parámetro buscado es

$$a = \frac{1}{\frac{\pi}{2}} = \frac{2}{\pi}.$$

§ 5. Sentido probabilístico de la función diferencial

Supongamos que $F(x)$ es la función integral de una magnitud aleatoria continua X. Por definición de función diferencial $f(x) = F'(x)$, o en otra forma

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

Como se sabe, la diferencia $F(x + \Delta x) - F(x)$ determina la probabilidad de que X tome un valor, perteneciente al intervalo $(x, x + \Delta x)$. En consecuencia, el límite de la relación de la probabilidad de que la magnitud aleatoria continua tome un valor perteneciente al intervalo $(x, x + \Delta x)$ a la longitud de este intervalo (para $\Delta x \to 0$), es igual al valor de la función diferencial en el punto x.

Por analogía con la definición de la densidad de masa en el punto*, el valor de la función $f(x)$ en el punto x conviene considerarlo como la densidad de probabilidad en ese punto.

De este modo, la función diferencial determina la densidad de distribución de probabilidad para cada punto x.

Del cálculo diferencial se sabe que el incremento de una función es aproximadamente igual a función diferencial, es decir,

$$F(x + \Delta x) - F(x) \simeq dF(x),$$

o bien

$$F(x + \Delta x) - F(x) \simeq F'(x)\, dx.$$

Puesto que $F'(x) = f(x)$ y $dx = \Delta x$, tendremos que

$$F(x + \Delta x) - F(x) \simeq f(x)\, \Delta x.$$

El sentido probabilístico de esta igualdad es que: la probabilidad de que una magnitud aleatoria tome un valor perteneciente al intervalo $(x, x + \Delta x)$ es aproximadamente igual (con la exactitud de hasta valores infinitamente pequeños de orden superior con respecto a Δx) al producto de la

---

* Si la masa está distribuida continuamente a lo largo del eje x por cierta ley, por ejemplo $F(x)$, la densidad $\rho(x)$ de masa en el punto x se llama límite de la relación entre la masa del intervalo $(x, x + \Delta x)$ y la longitud del intervalo cuando $\Delta x \to 0$, es decir $\rho(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$.

densidad de probabilidad en el punto x por la longitud del intervalo Δx.

Geométricamente este resultado se puede interpretar así: la probabilidad de que una magnitud aleatoria tome un valor perteneciente al intervalo $(x, x + \Delta x)$ es aproximadamente igual a la superficie del rectángulo de base Δx y altura $f(x)$.

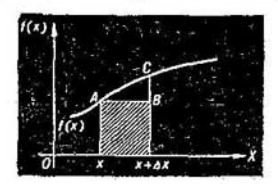


Fig. 5.

En la fig. 5 se aprecia que la superficie del rectángulo rayado es igual al producto $f(x)\,\Delta x$, sólo aproximadamente igual a la superficie del trapecio curvilíneo (probabilidad verdadera determinada por un intervalo definido $\int_x^{x+\Delta x} f(x)\,dx$). En este caso, el error admisible es igual a superficie del triángulo curvilíneo ABC.

§ 6. Ley de distribución uniforme de las probabilidades

Al resolver los problemas presentados por la práctica, se tropiezan con distintas distribuciones de magnitudes aleatorias continuas. Las funciones diferenciales de estas distribuciones se llaman también leyes de distribuciones. De ordinario se encuentran, por ejemplo, las leyes de distribuciones uniforme y normal. En este párrafo se examina la ley de distribución uniforme. El capítulo siguiente está dedicado a la ley de distribución normal.

La distribución de las probabilidades se llama *uniforme*, si en el intervalo, al que pertenecen todos los valores posibles de la magnitud aleatoria, la función diferencial tiene un valor constante.

Veamos un ejemplo de magnitud aleatoria continua uniformemente distribuida.

Ejemplo. La escala de un instrumento de medida está graduada en ciertas unidades. El error al redondear la lectura hasta una división entera próxima se puede considerar como una magnitud aleatoria X que puede tomar con la densidad constante de probabilidad cualquier valor entre dos divisiones enteras contiguas. Por consiguiente, X tiene distribución uniforme.

Hallamos la función diferencial de distribución uniforme, considerando que todos los valores posibles de la magnitud aleatoria acotados en el intervalo (a, b), en el que la función diferencial conserva un valor constante $f(x) = C$.



Fig. 6.

Según los datos X no toma valores fuera del intervalo (a, b), por eso $f(x) = 0$ cuando $x < a$ y $x > b$.

Hallamos el valor de C. Dado que todos los valores posibles de la magnitud aleatoria pertenecen al intervalo (a, b), debe cumplirse la igualdad

$$\int_a^b f(x)\,dx = 1, \text{ o bien } \int_a^b C\,dx = 1.$$

De donde

$$C = \frac{1}{\int_a^b dx} = \frac{1}{b-a}.$$

De este modo, la ley de distribución uniforme se puede escribir analíticamente, así:

$$f(x) = \begin{cases} 0 & \text{para} \quad x \leqslant a, \\ \frac{1}{b-a} & \text{para} \quad a < x \leqslant b, \\ 0 & \text{para} \quad x > b. \end{cases}$$

En la fig. 6 se muestra la gráfica de la función diferencial de distribución uniforme, mientras que en la fig. 4, la gráfica de la función integral.

Problemas

1. Una magnitud aleatoria se prefija por la función diferencial

$$f(x)=\begin{cases} 0 & \text{para} \quad x\leqslant -\dfrac{\pi}{2}; \\ a\cos x & \text{para} \quad -\dfrac{\pi}{2}<x\leqslant \dfrac{\pi}{2}; \\ 0 & \text{para} \quad x>\dfrac{\pi}{2}. \end{cases}$$

Hallar el coeficiente a.

Respuesta $a=\frac{1}{2}$.

2. Una magnitud aleatoria está dada por la función diferencial

$$f(x)=\begin{cases} 0 & \text{para} \quad x\leqslant 0; \\ \dfrac{1}{2}\operatorname{sen} x & \text{para} \quad 0<x\leqslant \pi; \\ 0 & \text{para} \quad x>\pi. \end{cases}$$

Hallar: a) la función integral; b) la probabilidad de que como resultado de la prueba la magnitud aleatoria toma un valor acotado en el intervalo $\left(0; \frac{\pi}{4}\right)$.

Respuesta a) $F(x)=\begin{cases} 0 & \text{para} \quad x\leqslant 0, \\ \dfrac{1}{2}(1-\cos x) & \text{para} \quad 0<x\leqslant \pi; \\ 1 & \text{para} \quad x>\pi; \end{cases}$

b) $\dfrac{2-\sqrt{2}}{4}$.

3. La magnitud aleatoria X se prefija por la función integral

$$F(x)=\begin{cases} 0 & \text{para} \quad x\leqslant 0, \\ x & \text{para} \quad 0<x\leqslant 1, \\ 1 & \text{para} \quad x>1. \end{cases}$$

Hallar la función diferencial.

Respuesta $f(x)=\begin{cases} 0 & \text{para} \quad x\leqslant 0. \\ 1 & \text{para} \quad 0<x\leqslant 1; \\ 0 & \text{para} \quad x>1. \end{cases}$

√4. La magnitud aleatoria X está prefijada por la función integral

$$F(x)=\begin{cases} 0 & \text{para} \quad x \leqslant 0, \\ \frac{1}{2}(1-\cos x) & \text{para} \quad 0 < x < \pi, \\ 1 & \text{para} \quad x > \pi. \end{cases}$$

Hallar la función diferencial.

Respuesta $f(x)=\begin{cases} 0 & \text{para} \quad x \leqslant 0, \\ \frac{1}{2}\operatorname{sen} x & \text{para} \quad 0 < x \leqslant \pi, \\ 0 & \text{para} \quad x > \pi. \end{cases}$

Capítulo doce

DISTRIBUCION NORMAL

§ 1. Características numéricas de las magnitudes aleatorias continuas

Extendemos las definiciones de las características numéricas de magnitudes discretas a las magnitudes continuas. Comencemos de la esperanza matemática.

Supongamos que una magnitud aleatoria continua X está prefijada por la función diferencial $f(x)$. Admitamos que todos los valores posibles de X pertenecen al segmento $[a, b]$. Descompongamos este segmento en n segmentos parciales de longitud $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$ y elijamos en cada uno de ellos un punto arbitrario x_i $(i = 1, 2, \ldots, n)$. Para determinar la esperanza matemática de una magnitud continua por semejanza con la discreta, formemos la suma de los productos de los valores posibles de x_i por sus probabilidades de caer en el intervalo Δx_i (recordemos que el producto $f(x)\,\Delta x$ es aproximadamente igual a la probabilidad de que X caiga en el intervalo Δx):

$$\sum x_i \cdot f(x_i)\,\Delta x_i.$$

Pasando al límite cuando la longitud del segmento máximo de los segmentos parciales tiende a cero, obtenemos la integral definida $\int_a^b xf(x)\,dx$.

Se llama *esperanza matemática de una magnitud aleatoria continua* X, cuyos valores posibles pertenecen al segmento $[a, b]$, la integral definida:

$$M(X) = \int_a^b xf(x)\,dx.$$

Si los valores posibles pertenecen a todo el eje x, tendremos que

$$M(x) = \int_{-\infty}^{\infty} xf(x)\,dx.$$

Se supone que la integral impropia converge absolutamente, es decir, existe la integral $\int_{-\infty}^{\infty} |x|\, f(x)\,dx$. Si esto no se cumpliese, el valor de la integral dependería de la velocidad de tendencia (por separado) del límite inferior hacia $-\infty$ y del superior hacia $+\infty$.

Por analogía con la dispersión de la magnitud aleatoria discreta se define también la de una magnitud continua.

Se llama *dispersión de una magnitud aleatoria continua* la esperanza matemática del cuadrado de su desviación.

Si los valores posibles de X pertenecen al segmento $[a, b]$, tendremos que

$$D(X) = \int_a^b [x - M(X)]^2 f(x)\,dx;$$

si los valores posibles corresponden a todo el eje x,

$$D(X) = \int_{-\infty}^{\infty} [x - M(x)]^2 f(x)\,dx.$$

La desviación cuadrática media de la magnitud aleatoria continua se determina al igual que para una magnitud discreta, por la igualdad

$$\sigma(X) = \sqrt{D(X)}.$$

Nota 1. Se puede demostrar que las propiedades de la esperanza matemática y de la dispersión de magnitudes discretas se extienden también a las magnitudes continuas.

Nota 2. Para el cálculo de la dispersión es fácil obtener fórmulas más convenientes:

$$D(X)=\int_a^b x^2 f(x)\,dx-[M(X)]^2,$$

$$D(X)=\int_{-\infty}^{\infty} x^2 f(x)\,dx-[M(X)]^2.$$

Ejemplo. Hallar la esperanza matemática y la dispersión de la magnitud aleatoria X prefijada por la función integral

$$F(x)=\begin{cases}0 & \text{para} \quad x\leqslant 0,\\ x & \text{para} \quad 0<x\leqslant 1,\\ 1 & \text{para} \quad x>1.\end{cases}$$

SOLUCION. Hallamos la función diferencial

$$f(x)=F'(x)=\begin{cases}0 & \text{para} \quad x\leqslant 0,\\ 1 & \text{para} \quad 0<x<1,\\ 0 & \text{para} \quad x>1.\end{cases}$$

Hallamos la esperanza matemática

$$M(X)=\int_0^1 x\cdot 1\cdot dx=\left.\frac{x^2}{2}\right|_0^1=\frac{1}{2}.$$

Hallamos la dispersión

$$D(X)=\int_0^1 x^2\cdot 1\cdot dx-\left[\frac{1}{2}\right]^2=\left.\frac{x^3}{3}\right|_0^1-\frac{1}{4}=\frac{1}{12}.$$

§ 2. Distribución normal

Se llama *normal* la distribución de las probabilidades de una magnitud aleatoria continua que se describe por la función diferencial

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Como podemos apreciar, *la distribución normal se determina por dos parámetros:* a y σ. Es suficiente prefijar estos parámetros para obtener la distribución normal. Mostremos

que el sentido probabilístico de estos parámetros es el siguiente: a es la esperanza matemática, σ, la desviación cuadrática media de la distribución normal.

a) Por definición de la esperanza matemática de una magnitud aleatoria continua

$$M(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx.$$

Introducimos un nuevo parámetro $z = \frac{x-a}{\sigma}$. De donde $x = \sigma z + a$, $dx = \sigma\,dz$. Teniendo en cuenta que los nuevos límites de integración son iguales a los anteriores, obtenemos

$$M(X) = \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + a)\, e^{-\frac{z^2}{2}}\,dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma z e^{-\frac{z^2}{2}}\,dz + \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz.$$

El primero de los sumandos es igual a cero (bajo el signo integral se halla una función impar; los límites de integración son simétricos respecto al origen de coordenadas). El segundo de los sumandos es igual a a $\left(\text{integral de Poisson } \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz = \sqrt{2\pi}\right)$.

De este modo, $M(X) = a$, es decir, *la esperanza matemática de la distribución normal es igual al parámetro a.*

b) Según definición de la dispersión de una magnitud aleatoria continua, tomando en consideración que $M(X) = a$, tendremos

$$D(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx.$$

Introducimos una nueva variable $z = \frac{x-a}{\sigma}$. De donde, $x - a = \sigma z$, $dx = \sigma dz$. Teniendo en cuenta que los nuevos límites de integración son iguales a los anteriores, obtenemos

$$D(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \cdot z e^{-\frac{z^2}{2}}\,dz.$$

Integrando por partes, poniendo $u = z$, $dv = ze^{-\frac{z^2}{2}}\, dz$, hallamos

$$D(X) = \sigma^2.$$

Por lo tanto,

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\sigma^2} = \sigma.$$

De esta manera, *la desviación cuadrática media de la distribución normal es igual al parámetro* σ.

Nota 1. La distribución normal de parámetros arbitrarios a y σ ($\sigma > 0$) se llama *general*.

La distribución normal de parámetros $a = 0$ y $\sigma = 1$ se llama *normada*. Por ejemplo, si X es una magnitud normal de parámetros a y σ, entonces $U = \frac{X-a}{\sigma}$ es una magnitud normal normada, además $M(U) = 0$, $\sigma(U) = 1$. La función diferencial de distribución normada

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Los valores de esta función están tabulados (suplemento 1).

Nota 2. La función integral de distribución normal general (cap. XI, § 3).

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(z-a)^2}{2\sigma^2}} dz,$$

y de la normada

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz.$$

Los valores de la función $F_0(x)$ están tabulados. Se comprueba fácilmente que

$$F(x) = F_0\left(\frac{x-a}{\sigma}\right).$$

Nota 3. La probabilidad de que una magnitud normal normada X caiga en el intervalo $(0, x)$ se puede hallar utilizando la función de Laplace $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$. En efecto (cap. XI, § 2), $P(0 <$

$$< X < x) = \int_0^x \varphi(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz = \Phi(x).$$

Nota 4. Tomando en consideración $\int\limits_{-\infty}^{\infty} \varphi(x)\,dx = 1$ (cap. XI, § 4), propiedad 2) y, por lo tanto, en virtud de la simetría de $\varphi(x)$ respecto a cero,

$\int\limits_{-\infty}^{0} \varphi(x)\,dx = 0{,}5$, por consiguiente, también $P(-\infty < X < 0) = 0{,}5$, es fácil obtener que $F_0(x) = 0{,}5 + \Phi(x)$.

En efecto, $F_0(x) = P(-\infty < X < x) = P(-\infty < X < 0) + P(0 < X < x) = 0{,}5 + \Phi(x)$.

§ 3. Curva normal

La gráfica de la función diferencial de distribución normal se llama *curva normal* (*curva de Gauss*).

Examinemos la función

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

por los métodos del cálculo diferencial.

1. Evidentemente, la función está definida en todo el eje x.

2. Para todos los valores de x la función toma valores positivos, es decir, la curva normal está situada encima del eje x.

3. El límite de la función, al crecer ilimitadamente x (en valor absoluto), es igual a cero: $\lim\limits_{|x|\to\infty} y = 0$, es decir, el eje x sirve de asíntota horizontal de la gráfica.

4. Examinemos la función respecto del extremo. Hallemos la derivada primera:

$$y' = -\frac{x-a}{\sigma^3\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Se aprecia fácilmente que $y' = 0$ cuando $x = a$, $y' > 0$ cuando $x < a$, $y' < 0$ para $x > a$. Por lo tanto, cuando $x = a$ la función tiene un máximo, igual a $\frac{1}{\sigma\sqrt{2\pi}}$.

5. La diferencia $x - a$ está contenida en la expresión analítica de la función al cuadrado, es decir, la gráfica de la función es simétrica respecto a la recta $x = a$.

6. Examinemos la función en el punto de inflexión.

Hallemos la derivada segunda

$$y'' = -\frac{1}{\sigma^3 \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \cdot \left[1 - \frac{(x-a)^2}{\sigma^2} \right].$$

Se nota fácilmente que para $x = a + \sigma$ y $x = a - \sigma$ la derivada segunda es igual a cero y, al pasar estos puntos,

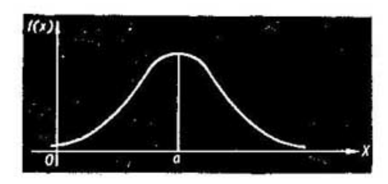


Fig. 7.

cambia de signo $\left(\text{en ambos puntos el valor de la función es igual a } \frac{1}{\sigma\sqrt{2\pi e}}\right)$. Por consiguiente, los puntos $\left(a-\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ y $\left(a+\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right)$ de la gráfica son puntos de inflexión.

En la fig. 7 se muestra la curva normal para $a = 1$; $\sigma = 2$.

§ 4. Influencia de los parámetros de la distribución normal sobre la fórmula de la curva normal

Aclaremos cómo influyen los valores de los parámetros a y σ en la fórmula y la disposición de la curva normal.

Se sabe que las gráficas de las funciones $f(x)$ y $f(x-a)$ tienen igual forma; desplazando la gráfica de $f(x)$ en el sentido positivo del eje x en a unidades de la escala para $a > 0$, o en sentido negativo para $a < 0$, obtenemos la gráfica de $f(x-a)$. De aquí se deduce que *la variación de la magnitud del parámetro a* (esperanza matemática) *no altera la forma de la curva normal, sino da lugar solamente a su desplazamiento a lo largo del eje x; hacia la derecha, si a crece, y hacia la izquierda, si a decrece.*

El problema cambia al variar el parámetro σ (desviación cuadrática media). Como se ha mostrado en el párrafo ante-

rior, el máximo de la función diferencial de distribución normal es igual a $\frac{1}{\sigma\sqrt{2\pi}}$. De donde se deduce que *al crecer σ, la ordenada máxima de la curva normal decrece, mientras que la propia curva deviene más suave, es decir, se aproxima al eje x; cuando decrece σ, la curva normal deviene más aguda y se alarga en sentido positivo del eje y.*

Cabe hacer notar que para todos los valores de los parámetros a y σ el área limitada por la curva normal y el eje x

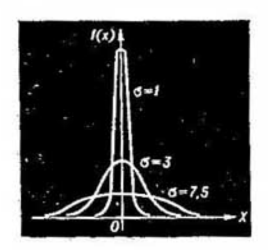


Fig. 8.

se mantiene igual a la unidad (cap. XI, § 4, segunda propiedad de la función diferencial).

En la fig. 8 están representadas las curvas normales para distintos valores de σ y $a = 0$. El dibujo ilustra claramente cómo la variación del parámetro σ influye en la forma de la curva normal.

Conviene señalar que cuando $a = 0$ y $\sigma = 1$ la curva normal $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ se llama *normada*.

§ 5. Probabilidad de que una magnitud aleatoria normal caiga en un intervalo dado

Ya sabemos que si una magnitud aleatoria X está dada por la función diferencial $f(x)$, la probabilidad de que X tome un valor perteneciente al intervalo (α, β), es la siguiente

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)\,dx.$$

Supongamos que la magnitud aleatoria X está distribuida por una ley normal. En tal caso, la probabilidad de que X tome un valor perteneciente al intervalo (α, β), es igual a

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx.$$

Transformamos esta fórmula de manera que se pueden utilizar las tablas hechas. Introducimos una nueva variable $z = \frac{x-a}{\sigma}$. De aquí, $x = \sigma z + a$, $dx = \sigma\, dz$. Hallamos los nuevos límites de integración. Si $x = \alpha$, tendremos que $z = \frac{\alpha-a}{\sigma}$; si $x = \beta$, $z = \frac{\beta-a}{\sigma}$.

Por consiguiente, tenemos

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-\frac{z^2}{2}} (\sigma\, dz) =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-a}{\sigma}}^{0} e^{-\frac{z^2}{2}}\, dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\beta-a}{\sigma}} e^{-\frac{z^2}{2}}\, dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\beta-a}{\sigma}} e^{-\frac{z^2}{2}}\, dz - \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{\alpha-a}{\sigma}} e^{-\frac{z^2}{2}}\, dz.$$

Utilizando la función de Laplace

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{z^2}{2}}\, dz,$$

finalmente obtenemos

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right). \qquad (*)$$

Ejemplo. Una magnitud aleatoria X está distribuida por una ley normal. La esperanza matemática y la desviación cuadrática media de esta magnitud son respectivamente iguales a 30 y 10. Hallar la probabilidad de que X tome un valor correspondiente al intervalo (10, 50).

SOLUCION. Utilizamos la fórmula (*). Según los datos del problema $\alpha = 10$, $\beta = 50$, $a = 30$, $\sigma = 10$, por lo tanto,

$$P(10 < X < 50) = \Phi\left(\frac{50-30}{10}\right) - \Phi\left(\frac{10-30}{10}\right) = 2\Phi(2).$$

Por la tabla (suplemento 2) hallamos

$$\Phi(2) = 0,4772.$$

De aquí, la probabilidad buscada es

$$P(10 < X < 50) = 2\cdot 0,4772 = 0,9544.$$

§ 6. Cálculo de la probabilidad de desviación prefijada

Frecuentemente se necesita calcular la probabilidad de que la desviación de una magnitud aleatoria normalmente distribuida X respecto del valor absoluto es menor que un número positivo dado a, es decir, hay que hallar la probabilidad de que se cumpla la desigualdad $|X - a| < \sigma$.

Sustituimos esta desigualdad por la doble desigualdad equivalente

$$-\delta < X - a < \delta,$$

o bien

$$a - \delta < X < a + \delta.$$

Utilizando la fórmula (*) (§ 5), obtenemos

$$P(|X-a| < \delta) = P(a-\delta < X < a+\delta) =$$

$$= \Phi\left[\frac{(a+\delta)-a}{\sigma}\right] - \Phi\left[\frac{(a-\delta)-a}{\sigma}\right] = \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right).$$

Tomando en consideración la igualdad

$$\Phi\left(-\frac{\delta}{\sigma}\right) = -\Phi\left(\frac{\delta}{\sigma}\right)$$

(la función de Laplace es impar), finalmente obtenemos

$$P(|X-a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

En particular, para $a = 0$

$$P(|X| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

En la fig. 9 se muestra claramente que, si dos magnitudes aleatorias están normalmente distribuidas y $a = 0$, la proba-

bilidad de tomar un valor, correspondiente al intervalo $(-\delta, \delta)$, es mayor para la magnitud que tiene un valor menor de σ. Este hecho corresponde totalemente al sentido

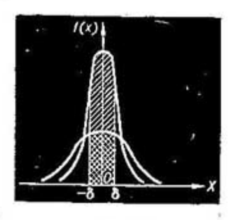


Fig. 9.

probabilístico del parámetro σ (σ es la desviación cuadrática media; este caracteriza la dispersión de la magnitud aleatoria alrededor de su esperanza matematica).

Nota. Evidentemente, los sucesos consistentes en el cumplimiento de las desigualdades $|X - a| < \delta$ y $|X - a| \geqslant \delta$ son opuestos. Por eso, si la probabilidad de que se cumpla la desigualdad $|X - a| < \delta$ es igual a p, la probabilidad de la desigualdad $|X - a| \geqslant \delta$ es igual a $1 - p$.

Ejemplo. La magnitud aleatoria X está distribuida normalmente. La esperanza matemática y la desviación cuadrática media de X son respectivamente iguales a 20 y 10. Hallar la probabilidad de que la desviación sea en valor absoluto menor que tres.

SOLUCION. Utilizamos la fórmula

$$P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Según los datos del problema $\delta = 3$, $a = 20$, $\sigma = 10$. Por lo tanto,

$$P(|X-20|<3)=2\Phi\left(\frac{3}{10}\right)=2\Phi\ (0,3).$$

Por la tabla (suplemento 2) hallamos $\Phi(0,3) = 0,1179$. La probabilidad buscada es

$$P(|X-20|<3) = 0,2358.$$

§ 7. Regla de las tres sigmas

Transformamos la fórmula (§ 6)

$$P(|X-a|<\delta)=2\Phi\left(\frac{\delta}{\sigma}\right),$$

poniendo $\delta=\sigma t$. En conclusión obtenemos,

$$P(|X-a|<\sigma t)=2\Phi(t).$$

Si $t=3$ y, por lo tanto, $\sigma t=3\sigma$, tendremos que

$$P(|X-a|<3\sigma)=2\Phi(3)=2\cdot 0{,}49865=0{,}9973,$$

es decir, la probabilidad de que la desviación sea en valor absoluto menor que el triple de la desviación cuadrática media, es igual a 0,9973.

En otras palabras, la probabilidad de que la magnitud absoluta de la desviación sea *mayor* que el triple de la desviación cuadrática media, es muy pequeña y, precisamente, igual a 0,0027. Esto significa que sólo en 0,27% de los casos puede ocurrir así. Estos sucesos, partiendo del principio de imposiblidad de los sucesos poco probables, pueden considerarse prácticamente inciertos o imposibles. En esto reside precisamente la esencia de la regla de las tres sigmas: *si una magnitud aleatoria está distribuida normalmente, la magnitud absoluta de su desviación respecto de la esperanza matemática no es mayor que el triple de la desviación cuadrática media*

En la práctica, la regla de las tres sigmas se aplica así: si la distribución de la magnitud aleatoria que se estudia no se conoce, pero si se cumple la condición indicada en la regla expuesta, se puede suponer que la magnitud a estudiar está distribuida normalmente; en caso contrario no está distribuida normalmente.

§ 8. Noción del teorema de Liapunov

Se sabe que las magnitudes aleatorias normalmente distribuidas están profusamente difundidas en la práctica. ¿A qué se debe esto? La respuesta la dio el eminente matemático ruso A. M. Liapunov (teorema límite central de la teoría de las probabilidades). Damos sólo el *corolario del teorema de Liapunov: si una magnitud aleatoria X es la suma de un número muy grande de magnitudes aleatorias mutuamente indepen-*

dientes, la influencia de cada una de ellas en toda la suma es despreciable, entonces X tiene una distribución próxima a la normal.

En la práctica se tropieza con mayor frecuencia, precisamente, con estas magnitudes aleatorias.

El ejemplo siguiente aclara lo dicho.

Ejemplo. Supongamos que se mide cierta magnitud física. Cualquier medición da solamente un valor aproximado de la magnitud a medir, puesto que sobre el resultado de la medición influyen muchos factores aleatorios independientes (temperatura, oscilación del instrumento, humedad, etc.). Cada uno de estos factores engendra un «error parcial» ínfimo. Sin embargo, dado que el número de estos factores es muy grande, el conjunto de su acción ocasiona ya un «error total» notable.

Considerando el error total como suma de un número muy grande de errores particulares mutuamente independientes, podemos deducir que el error total tiene una distribución próxima a la normal. La experiencia confirma la validez de esta deducción.

§ 9. Estimación de la desviación de la distribución teórica de la normal. Asimetría y exceso

La distribución de frecuencias relativas se llama *empírica*. La estadística matemática estudia la distribución empírica.

La distribución de las probabilidades se llama *teórica*. La distribución teórica es estudiada por la teoría de las probabilidades. En este párrafo se examinan las distribuciones teóricas.

Al estudiar las distribuciones, diferentes de la normal, surge la necesidad de estimar cuantitativamente esta diferencia. Con este propósito se introducen características especiales, en particular, la asimetría y el exceso. Para la distribución normal estas características son iguales a cero. Por eso, si para la distribución a estudiar la asimetría y el exceso tienen pequeños valores, se puede suponer la proximidad de esta distribución a la normal. Por el contrario, los grandes valores de la asimetría y del exceso denotan la desviación considerable de la normal.

¿Cómo estimar la asimetría? Se puede demostrar que, para la distribución simétrica (la gráfica de tal distribución

es simétrica con respecto a la recta $x = M(X)$) cada momento central de orden impar es igual a cero. Para las distribuciones asimétricas los momentos centrales de orden impar son distintos de cero. Por lo cual, cualquiera de estos momentos (fuera del momento de primer orden que es igual a cero para toda distribución) puede servir para estimar la asimetría; es natural elegir el más simple de ellos, es decir, el momento de tercer orden μ_3. Sin embargo, no es conveniente tomar este momento para estimar la asimetría ya que su magnitud depende de las unidades, con las que se mide la magnitud aleatoria. Para evitar este inconveniente, μ_3 se divide por σ^3 y, de este modo, se obtiene una característica adimensional.

Se llama *asimetría de la distribución teórica* la relación entre el momento central de tercer orden y el cubo de la desviación cuadrática media:

$$A_s = \frac{\mu_3}{\sigma^3} ,$$

La simetría es positiva, si la «parte larga» de la curva de distribución se encuentra a la derecha de la esperanza matemática; la asimetría es negativa, si la «parte larga» de la curva se encuentra a la izquierda de la esperanza matemática. El signo de la asimetría se determina prácticamente por la ubicación de la curva de distribución con respecto a la moda (punto máximo de la función diferencial): si la parte alargada de la curva de distribución se encuentra a la derecha de la moda, la asimetría es positiva (fig. 10, *a*), si se encuentra a la izquierda, negativa (fig. 10, *b*).

Para estimar el «declive», es decir, la pendiente máxima o mínima de la curva de distribución teórica en comparación con la curva normal se utiliza la característica exceso.

Se llama *exceso de la distribución teórica* la característica determinada por la igualdad

$$E_k = \frac{\mu_4}{\sigma^4} - 3.$$

Para la distribución normal $\frac{\mu_4}{\sigma^4} = 3$, y, por lo tanto, el exceso es igual a cero. Por eso, si el exceso de cierta distribución es distinto de cero, la curva de esta distribución se diferencia de la curva normal: si el exceso es positivo, la curva tiene un vértice más alto y «agudo» que la curva

normal (fig. 11, *a*); si el exceso es negativo, la curva a comparar tiene un vértice más bajo y «plano» que la curva nor-

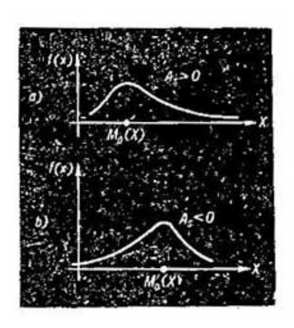


Fig. 10

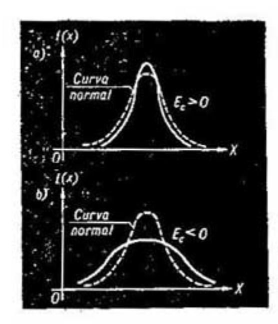


Fig. 11

mal (fig. 11, *b*). En este caso se supone que las distribuciones normal y teórica tienen iguales las esperanzas matemáticas y las dispersiones.

§ 10. Función de un argumento aleatorio y su distribución

Previamente cabe hacer notar que, en adelante, en vez de decir «ley de distribución de las probabilidades», diremos simplemente «distribución».

Si a cada valor posible de una magnitud aleatoria X corresponde un valor posible de la magnitud aleatoria Y, entonces Y se llama función del argumento aleatorio X:

$$Y = \varphi(X).$$

Más adelante se muestra como hallar la distribución de la función por la distribución conocida del argumento discreto y continuo.

1. *Supongamos que el argumento X es una magnitud aleatoria discreta.*

a) Si a los distintos valores posibles del argumento X corresponden *distintos* valores posibles de la función Y,

las probabilidades de los correspondientes valores de X e Y son iguales entre sí.

Ejemplo 1. La magnitud aleatoria discreta X está prefijada por la distribución:

X	2	3
p	0,6	0,4

Hallar la distribución de la función $Y = X^2$.

SOLUCION. Hallamos los valores posibles de Y:

$$y_1 = 2^2 = 4; \quad y_2 = 3^2 = 9.$$

Escribimos la distribución buscada de Y:

Y	4	9
p	0,6	0,4.

b) Si a los distintos valores posibles de X corresponden los valores de Y, entre los cuales hay iguales entre sí, entonces hay que adicionar las probabilidades de los valores que se repiten de Y.

Ejemplo 2. La magnitud aleatoria discreta X se prefija por la distribución:

X	-2	2	3
p	0,4	0,5	0,1.

Hallar la distribución de la función $Y = X^2$.

SOLUCION. La probabilidad del valor posible de $y_1 = 4$ es igual a la suma de las probabilidades de los sucesos mutuamente excluyentes $X = -2$, $X = 2$, es decir, $0,4 + 0,5 = 0,9$. La probabilidad del valor posible de $y_2 = 9$ es igual a 0,1. Escribimos la distribución buscada de Y:

Y	4	9
p	0,9	0,1.

2. *Supongamos que el argumento X es una magnitud aleatoria continua.* ¿Cómo hallar la distribución de la función $Y = \varphi(X)$, conociendo la función diferencial $f(x)$ del argumento aleatorio X? Se demostró que: si $y = \varphi(x)$ es una función diferenciable rigurosamente creciente o rigurosamente decreciente, cuya función inversa $x = \psi(y)$, la función diferencial $g(y)$ de la magnitud aleatoria Y se halla por la igualdad

$$g(y) = f[\psi(y)] \cdot |\psi'(y)|.$$

Ejemplo 3. La magnitud aleatoria X está distribuida normalmente, además su esperanza matemática es $a = 0$. Hallar la distribución de la función $Y = X^3$.

SOLUCION. Puesto que la función $y = x^3$, es diferenciable y rigurosamente creciente, se puede utilizar la fórmula:

$$g(y) = f[\psi(y)] \cdot |\psi'(y)|. \qquad (*)$$

Hallamos la función, inversa a la función $y = x^3$:

$$\psi(y) = x = y^{\frac{1}{3}}.$$

Hallamos $f[\psi(y)]$. Por los datos

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

por eso

$$f[\psi(y)] = f[y^{\frac{1}{3}}] = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^{\frac{2}{3}}}{2\sigma^2}}. \qquad (**)$$

Hallamos la derivada de la función inversa de y:

$$\psi'(y) = (y^{\frac{1}{3}})' = \frac{1}{3y^{\frac{2}{3}}}. \qquad (***)$$

Hallamos la función diferencial buscada, para lo cual ponemos (**) y (***) en (*):

$$g(y) = \frac{1}{3\sigma y^{\frac{2}{3}}\sqrt{2\pi}} e^{-\frac{y^{\frac{2}{3}}}{2\sigma^2}}.$$

Nota. Utilizando la fórmula (*) se puede demostrar que la función lineal $Y = AX + B$ de argumento X normalmente distribuido también está normalmente distribuida, además para hallar la esperanza matemática de Y, hay que poner en la expresión de la función en lugar del argumento X, su esperanza matemática a:

$$M(Y) = Aa + B;$$

para hallar la desviación cuadrática media de Y, hay que multiplicar la desviación cuadrática media del argumento X por el módulo del coeficiente de X:

$$\sigma(Y) = |A| \cdot \sigma(X).$$

Ejemplo 4. Hallar la función diferencial de la función lineal $Y = 3X + 1$, si el argumento está distribuido normalmente, además, la esperanza matemática de X es igual a 2 y la desviación cuadrática media es igual a 0,5.

SOLUCION. Hallar la esperanza matemática de Y

$$M(Y) = 3 \cdot 2 + 1 = 7.$$

Hallamos la desviación cuadrática media de Y

$$\sigma(Y) = 3 \cdot 0{,}5 = 1{,}5.$$

La función diferencial buscada tiene la forma

$$g(y) = \frac{1}{1{,}5\sqrt{2\pi}} e^{-\frac{(y-7)^2}{2\cdot(1{,}5)^2}}.$$

§ 11. Esperanza matemática de la función de un argumento aleatorio

Dada la función $Y = \varphi(X)$ del argumento aleatorio X, hallar la esperanza matemática de esta función, conociendo la ley de distribución del argumento.

1. *Supongamos que el argumento X es una magnitud aleatoria discreta* con los valores posibles $x_1, x_2, \ldots, x_n$, cuyas probabilidades son respectivamente iguales *a* p_1, $p_2, \ldots, p_n$. Evidentemente, Y también es una magnitud aleatoria discreta con los valores posibles $y_1 = \varphi(x_1)$, $y_2 = \varphi(x_2), \ldots \varphi y_n = \varphi(x_n)$. Ya que el suceso «la magnitud X tomó el valor x_i» da lugar al suceso «la magnitud Y tomó el valor $\varphi(x_i)$», las probabilidades de los valores posibles de Y son respectivamente iguales a $p_1, p_2, \ldots, p_n$. Por lo tanto, la esperanza matemática de la función es

$$M[\varphi(X)] = \sum_{i=1}^{n} \varphi(x_i) \cdot p_i. \qquad (*)$$

Ejemplo 1. La magnitud aleatoria discreta X está dada por la distribución

X	1	3	5
p	0,2	0,5	0,3.

Hallar la esperanza matemática de la función $Y = \varphi(X) = X^2 + 1$.

SOLUCION. Hallamos los valores posibles de Y:

$$\varphi(1) = 1^2 + 1 = 2; \quad \varphi(3) = 3^2 + 1 = 10;$$
$$\varphi(5) = 5^2 + 1 = 26.$$

La esperanza matemática buscada de la función es

$$M[X^2+1] = 2\cdot 0{,}2 + 10\cdot 0{,}5 + 26\cdot 0{,}3 = 13{,}2.$$

2. *Supongamos que el argumento X es una magnitud aleatoria continua prefijada* por la función diferencial $f(x)$. Para hallar la esperanza matemática de la función $Y = \varphi(X)$ al principio se puede hallar la función $g(y)$ de la magnitud Y y, luego, utilizar la fórmula

$$M(Y) = \int_{-\infty}^{\infty} y\cdot g(y)\,dy.$$

Sin embargo, si la función diferencial $g(y)$ es difícil de hallar, se puede hallar directamente la esperanza matemática de la función $\varphi(X)$ por la fórmula

$$M[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f(x)\,dx.$$

En particular, si los valores posibles de X pertenecen al intervalo (a, b),

$$M[\varphi(X)] = \int_{a}^{b} \varphi(x) f(x)\,dx. \qquad (^{**})$$

Omitiendo la demostración, cabe hacer notar que ella es análoga a la demostración de la fórmula (*), si se sustituye la suma por la integración, y la probabilidad, por el elemento de la probabilidad $f(x)\,\Delta x$.

Ejemplo 2. La magnitud aleatoria continua X está prefijada por la función diferencial $f(x) = \operatorname{sen} x$ en el intervalo $(0; \frac{\pi}{2})$; $f(x) = 0$ fuera de este intervalo. Hallar la esperanza matemática de la función $Y = \varphi(X) = X^2$.

SOLUCION. Utilizamos la fórmula (**). Por los datos del problema $f(x) = \operatorname{sen} x$, $\varphi(x) = x^2$, $a = 0$, $b = \frac{\pi}{2}$.

Por lo tanto,

$$M[\varphi(x)] = \int_{0}^{\frac{\pi}{2}} x^2 \operatorname{sen} x\,dx.$$

Integrando por partes, obtenemos la esperanza matemática buscada

$$M[X^2] = \pi - 2.$$

§ 12. Función de dos argumentos aleatorios. Distribución de la suma de sumandos independientes. Estabilidad de distribución normal

Si a cada par de valores posibles de las magnitudes aleatorias X e Y le corresponde un valor posible de la magnitud aleatoria Z, entonces Z se llama *función de dos argumentos aleatorios* X e Y:

$$Z = \varphi(X, Y).$$

A continuación se mostrará mediante ejemplos cómo hallar la distribución de la función

$$Z = X + Y$$

conociendo las distribuciones de los sumandos. Semejante problema se encuentra frecuentemente en la práctica. Por ejemplo, si X es el error de lecturas del instrumento de medida (distribuido normalmente), Y es el error de redondeo de las lecturas hasta la división próxima de la escala (distribuido normalmente), surge el problema de hallar la ley de distribución de la suma de los errores $Z = X + Y$.

1. *Supongamos que X e Y son magnitudes aleatorias independientes discretas.* Para componer la ley de distribución de la función $Z = X + Y$, hay que hallar todos los valores posibles de Z y sus probabilidades.

Ejemplo 1. Las magnitudes aleatorias independientes discretas están prefijadas por las distribuciones:

X	1	2	Y	3	4
p	0,4	0,6	p	0,2	0,8

Componer la distribución de la magnitud aleatoria $Z = X + Y$.

SOLUCION. Los valores posibles de Z son las sumas de cada valor posible de X con todos los valores posibles de Y:

$$z_1 = 1 + 3 = 4; \quad z_2 = 1 + 4 = 5; \quad z_3 = 2 + 3 = 5;$$
$$z_4 = 2 + 4 = 6.$$

Hallamos las probabilidades de estos valores posibles Para que $Z = 4$, es suficiente que la magnitud X tome el valor $x_1 = 1$ y la magnitud Y, el valor $y_1 = 3$. Las probabilidades de estos valores posibles, como se deduce de las

leyes de distribuciones dadas, son respectivamente iguales a 0,4 y 0,2.

Puesto que los argumentos X e Y son independientes, los sucesos $X = 1$ e $Y = 3$ son independientes, y, por lo tanto, la probabilidad de que ocurran simultáneamente (es decir, la probabilidad del suceso $Z = 1 + 3 = 4$) por el teorema del producto es igual a $0{,}4 \cdot 0{,}2 = 0{,}08$.

Análogamente hallamos:

$$P\,(Z = 1 + 4 = 5) = 0{,}4 \cdot 0{,}8 = 0{,}32;$$
$$P\,(Z = 2 + 3 = 5) = 0{,}6 \cdot 0{,}2 = 0{,}12;$$
$$P\,(Z = 2 + 4 = 6) = 0{,}6 \cdot 0{,}8 = 0{,}48.$$

Escribimos la distribución buscada, sumando previamente las probabilidades de los sucesos mutuamente excluyentes $Z = z_2$, $Z = z_2$ $(0{,}32 + 0{,}12 = 0{,}44)$:

Z	4	5	6
p	0,08	0,44	0,48.

Verificación: $0{,}08 + 0{,}44 + 0{,}48 = 1$.

2. *Supongamos que X e Y son magnitudes aleatorias continuas.* Se demostró que: si X e Y son independientes, la función diferencial $g\,(z)$ de la suma $Z = X + Y$ (a condición de que la función diferencial por lo menos de uno de los argumentos se encuentre en el intervalo $(-\infty, \infty)$ por una fórmula) puede hallarse por la igualdad

$$g\,(z) = \int_{-\infty}^{\infty} f_1\,(x)\, f_2\,(z - x)\, dx \qquad (^*)$$

o bien por la igualdad equivalente

$$g\,(z) = \int_{-\infty}^{\infty} f_1\,(z - y)\, f_2\,(y)\, dy, \qquad (^{**})$$

donde f_1, f_2 son las funciones diferenciales de los argumentos.

Si los valores posibles de los argumentos no son negativos, $g\,(z)$ se hallan por la fórmula

$$g\,(z) = \int_{0}^{z} f_1\,(x)\, f_2\,(z - x)\, dx, \qquad (^{***})$$

o bien por la fórmula equivalente

$$g(z)=\int_0^z f_1(z-y)\, f_2(y)\, dy. \qquad (****)$$

La función diferencial de la suma de magnitudes aleatorias independientes se llama *composición.*

La ley de distribución de las probabilidades se llama *estable*, si la composición de tales leyes es la misma ley (que se diferencia, en general, por los parámetros). La ley normal posee la propiedad de estabilidad: la composición de leyes normales también tiene distribución normal (la esperanza matemática y la dispersión de esta composición son iguales respectivamente a las sumas de las esperanzas matemáticas y las dispersiones de los sumandos). Por ejemplo, si X e Y son magnitudes aleatorias independientes distribuidas normalmente con esperanzas matemáticas y dispersiones respectivamente iguales a $a_1 = 3$, $a_2 = 4$, $D_1 = 1$, $D_2 = 0{,}5$, la composición de estas magnitudes (es decir, la función diferencial de la suma $Z = X + Y$) también está distribuida normalmente; además, la esperanza matemática y la dispersión de la composición son respectivamente iguales a $a = 3 + 4 = 7$; $D = 1 + 0.5 = 1{,}5$.

Ejemplo 2. Las magnitudes aleatorias independientes X e Y están prefijadas por las funciones diferenciales:

$$f(x)=\frac{1}{3}e^{-\frac{x}{3}} \qquad (0 \leqslant x < \infty);$$

$$f(y)=\frac{1}{4}e^{-\frac{y}{4}} \qquad (0 \leqslant y < \infty).$$

Hallar la composición de estas leyes, es decir, la función diferencial de la magnitud aleatoria $Z = X + Y$.

SOLUCION. Ya que los valores posibles de los argumentos no son negativos, utilizamos la fórmula (***)

$$g(z)=\int_0^z f_1(x)\, f_2(z-x)\, dx=\int_0^z \left[\frac{1}{3}e^{-\frac{x}{3}}\right]\left[\frac{1}{4}e^{-\frac{z-x}{4}}\right]dx=$$

$$=\frac{1}{12}e^{-\frac{z}{4}}\int_0^z e^{-\frac{x}{12}}\,dx=e^{-\frac{z}{4}}\left(1-e^{-\frac{z}{12}}\right).$$

Cabe hacer notar que aquí $z \geqslant 0$, ya que $Z = X + Y$ y por los datos los valores posibles de X e Y no son negativos.

Recomendamos al lector cerciorarse de que

$$\int_0^\infty g(z)\,dz = 1.$$

En los siguientes párrafos se describen brevemente las distribuciones vinculadas con la normal, que se utilizaran al exponer la estadística matemática.

§ 13. Distribución χ^2

Supongamos que X_i $(i = 1, 2, \ldots, n)$ son magnitudes aleatorias independientes normales; además, la esperanza matemática de cada una de ellas es igual a cero, y la desviación cuadrática media, igual a la unidad. En tal caso, la suma de los cuadrados de estas magnitudes

$$x^2 = \sum_{i=1}^{n} X_i^2$$

está distribuida por la ley χ^2 («ji cuadrado») con $k = n$ grados de libertad; si estas magnitudes están vinculadas por una correlación lineal, por ejemplo $\sum X_i = n\overline{X}$, el número de grados de libertad es $k = n - 1$.

La función diferencial de esta distribución es

$$f(x) = \begin{cases} 0 & \text{para } x \leqslant 0, \\ \dfrac{1}{2^{\frac{k}{2}} \cdot \Gamma\left(\frac{k}{2}\right)} e^{-\frac{x}{2}} x^{\frac{k}{2}-1} & \text{para } x > 0, \end{cases}$$

donde $\Gamma(x) \int_0^\infty t^{x-1} e^{-t}\,dt$ es la función gamma; en particular, $\Gamma(n + 1) = n!$

De aquí se aprecia que la distribución «ji cuadrado» se determina por un parámetro, o sea, el número de grados de libertad k.

Al aumentar el número de grados de libertad, la distribución se aproxima lentamente a la normal.

§ 14. Distribución t de Student

Supongamos que Z es una magnitud aleatoria normal; además, $M(Z) = 0$, $\sigma(Z) = 1$, y V es una magnitud independiente de Z, distribuida por la ley χ^2 con k grados de libertad. En tal caso, la magnitud

$$T = \frac{Z}{\sqrt{\frac{V}{k}}} \qquad (*)$$

tiene una distribución llamada distribución t o distribución t de Student (seudónimo del estadístico inglés V. Gosset) con k grados de libertad.

De esta manera, la relación entre la magnitud normal normada y la raíz cuadrada de la magnitud aleatoria independiente, distribuida por la ley «ji cuadrado» con k grados de libertad, dividida por k, se distribuye según la ley de Student con k grados de libertad.

Al crecer el número de grados de libertad la distribución t de Student se aproxima rápidamente a la normal. Más adelante se dan conocimientos suplementarios sobre esta distribución (cap. XVI, § 16).

§ 15. Distribución F de Fisher — Snedecor

Si U y V son magnitudes aleatorias independientes, distribuidas por la ley χ^2 con grados de libertad k_1 y k_2 respectivamente, la magnitud

$$F = \frac{\frac{U}{k_1}}{\frac{V}{k_2}} \qquad (*)$$

tiene una distribución llamada distribución F de Fischer-Snedecor con grados de libertad k_1 y k_2 (a veces se la designa por V^2). La función diferecial es

$$f(x) = \begin{cases} 0 & \text{para } x \leqslant 0, \\ C_0 \dfrac{x^{\frac{k_1-2}{2}}}{(k_2 + k_1 x)^{\frac{k_1+k_2}{2}}} & \text{para } x > 0, \end{cases}$$

donde $C_0 = \dfrac{\Gamma\left(\frac{k_1+k_2}{2}\right) \cdot k_1^{\frac{k_2}{2}} \cdot k_2^{\frac{k_2}{2}}}{\Gamma\left(\frac{k_1}{2}\right) \Gamma\left(\frac{k_2}{2}\right)}$.

Podemos apreciar que la distribución F se determina por dos parámetros, o sea, los números de grados de libertad. Más adelante se dan conocimientos suplementarios sobre esta distribución (cap. XIX, § 8).

Problemas

1. Hallar la esperanza matemática y la dispersión de la magnitud aleatoria X, conociendo su función diferencial: a) $f(x) = \frac{1}{\pi\sqrt{1-x^2}}$ para $-1 < x < 1$, $f(x) = 0$ para los demás valores de x;

b) $f(x) = \frac{1}{2 \cdot l}$ para $a - l \leqslant x \leqslant a + l$, $f(x) = 0$ para los demás valores de x.

Respuesta a) $M(X) = 0$, $D(X) = \frac{1}{2}$; b) $M(X) = a$, $D(X) = \frac{l^2}{3}$

2. La magnitud aleatoria X está distribuida normalmente. La esperanza matemática y la desviación cuadrática media de esta magnitud son respectivamente iguales a 6 y 2. Hallar la probabilidad de que como resultado de lexperimento X toma un valor acotado en el intervalo (4; 8).

Respuesta 0,6826.

3. Una magnitud aleatoria está distribuida normalmente. La desviación cuadrática media de esta magnitud es igual a 0,4. Hallar la probabilidad de que la desviación de la magnitud aleatoria respecto de su esperanza matemática será en valor absoluto menor de 0,3.

Respuesta 0,5468.

4. Los errores aleatorios de medición obedecen a una ley normal con una desviación cuadrática media $\sigma = 1$ mm y la esperanza matemática $a = 0$. Hallar la probabilidad de que de dos observaciones independientes el error por lo menos de uno de ellos no supera en valor absoluto la magnitud 1,28 mm.

Respuesta 0,96.

5. Los ejes, producidos por el automático, se consideran standards si la desviación del diámetro del eje de las dimensiones proyectadas no es mayor de 2 mm. Las desviaciones aleatorias del diámetro de los ejes obedecen a una ley normal con una desviación cuadrática media $\sigma = 1,6$ mm y esperanza matemática $a = 0$. ¿Qué porcentaje de ejes standards produce el automático?

Respuesta. Aproximadamente 79%.

6. La magnitud aleatoria discreta está dada por la ley de distribución

a)

X	1	2	3
p	0,2	0,1	0,7;

b)

X	-1	1	2
p	0,1	0,2	0,7

Hallar la ley de distribución de la magnitud aleatoria $Y = X^4$.

Respuesta a)

Y	1	16	81
p	0,2	0,1	0,7;

b)

Y	-1	16
p	0,3	0,7.

7. La magnitud aleatoria continua está dada por la función diferencial $f(x)$. Hallar la función diferencial $g(y)$ de la magnitud aleatoria Y, si:
a) $Y = X + 1 \quad (-\infty < x < \infty)$;
b) $Y = 2X \quad (-a < x < a)$.

Respuesta a) $g(y) = f(y-1) \quad (-\infty < y < \infty)$;

b) $g(y) = \frac{1}{2} f\left(\frac{y}{2}\right) \quad (-2a < y < 2a)$.

8. Las magnitudes aleatorias discretas independientes están prefijadas por las siguientes leyes de distribución:

X	2	3	5
p	0,3	0,5	0,2;

Y	1	4
p	0,2	0,8

Hallar las leyes de distribución de las funciones: a) $Z = X + Y$; b) $Z = XY$.

Respuesta a)

Z	3	4	6	7	9
p	0,06	0,10	0,28	0,40	0,16;

b)

Z	2	3	5	8	12	20
p	0,06	0,10	0,04	0,24	0,40	0,16.

9. Las magnitudes aleatorias independientes X e Y están dadas por las funciones diferenciales:

$$f_1(x) = \frac{1}{3} e^{-\frac{x}{3}} \qquad (0 \leqslant x < \infty);$$

$$f_2(x) = \frac{1}{5} e^{-\frac{y}{5}} \qquad (0 \leqslant y < \infty).$$

Hallar la composición de estas leyes, es decir, la función diferencial de la magnitud aleatoria $Z = X + Y$.

Respuesta $$g(z) = \begin{cases} \frac{1}{2} e^{-\frac{z}{5}} \left(1 - e^{-\frac{2z}{15}}\right) & \text{para } z \geqslant 0; \\ 0 & \text{para } z < 0. \end{cases}$$

DISTRIBUCION EXPONENCIAL

§ 1. Definición de la distribución exponencial

Se llama *exponencial* la distribución de las probabilidades de una magnitud aleatoria continua X que se describe por la función diferencial $f(x)=\begin{cases} 0 & \text{para } x<0, \\ \lambda^{-\lambda x} & \text{para } x\geqslant 0, \end{cases}$
donde λ es una magnitud constante positiva.

Podemos apreciar que la distribución exponencial se determina por *un* parámetro λ. Esta particularidad de la distribución exponencial indica su ventaja en comparación con las distribuciones dependientes de muchos parámetros. Generalmente, los parámetros no se conocen y hay que hallar sus estimaciones (valores aproximados); pero, es más fácil estimar un parámetro que dos, o tres, etc

El tiempo *entre* las apariciones de dos sucesos consecutivos de flujo elemental son de ejemplo de una magnitud aleatoria continua distribuida por una ley exponencial (véase § 5).

Hallamos la función integral de la distribución exponencial (cap. XI, § 3).

$$F(x)=\int_{-\infty}^{x} f(x)\,dx=\int_{-\infty}^{0} 0\cdot dx+\lambda\int_{0}^{x} e^{-\lambda x}\,dx=1-e^{-\lambda x}.$$

De este modo, $F(x)=\begin{cases} 0 & \text{para } x<0, \\ 1-e^{-\lambda x} & \text{para } x\geqslant 0. \end{cases}$

Hemos determinado la distribución exponencial mediante la función diferencial; está claro que se puede determinar también utilizando la función integral.

En la fig. 12 se muestran las gráficas de las funciones diferencial e integral de la distribución exponencial.

Ejemplo. Escribir las funciones diferencial e integral de la distribución exponencial, si el parámetro $\lambda=8$.

SOLUCION. Evidentemente,

$$f(x)=8e^{-8x} \text{ para } x\geqslant 0; \quad f(x)=0 \text{ para } x<0.$$
$$F(x)=1-e^{-8x}.$$

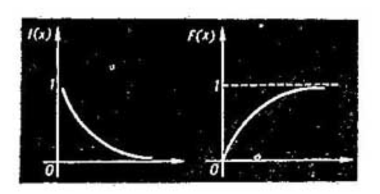


Fig. 12.

§ 2. Probabilidad de que una magnitud aleatoria distribuida de modo exponencial caiga en un intervalo dado

Hallemos la probabilidad de que la magnitud aleatoria continua X distribuida por la ley exponencial que está prefijada por la función integral

$$F(x) = 1 - e^{-\lambda x},$$

caiga en el intervalo (a, b).

Utilizamos la fórmula (cap. X, § 2, corolario 1)

$$P(a < X < b) = F(b) - F(a).$$

Tomando en consideración que $F(a) = 1 - e^{-\lambda a}$, $F(b) = 1 - e^{-\lambda b}$, obtenemos

$$P(a < X < b) = e^{-\lambda a} - e^{-\lambda b}. \qquad (*)$$

Los valores de la función e^{-x} se hallan por la tabla.

Ejemplo. La magnitud aleatoria continua X está distribuida según la ley exponencial:

$$f(x) = 2e^{-2x} \text{ para } x \geqslant 0; \; f(x) = 0 \text{ para } x < 0.$$

Hallar la probabilidad de que como resultado del experimento X cae en el intervalo $(0,3; 1)$.

SOLUCION. Por los datos $\lambda = 2$. Utilizamos la fórmula (*):

$$P(0,3 < X < 1) = e^{-(2\cdot 0,3)} - e^{-(2\cdot 1)} = e^{-0,6} - e^{-2} =$$
$$= 0,54881 - 0,13534 \simeq 0,41.$$

§ 3. Características numéricas de la distribución exponencial

Supongamos que la magnitud aleatoria continua X está distribuida por la ley exponencial

$$f(x)=\begin{cases}0 & \text{para } x<0,\\ \lambda e^{-\lambda x} & \text{para } x\geqslant 0.\end{cases}$$

Hallamos la esperanza matemática (cap. XII, § 1):

$$M(X)=\int_0^\infty x f(x)\,dx=\lambda\int_0^\infty x e^{-\lambda x}\,dx.$$

Integrando por partes, obtenemos

$$M(X)=\frac{1}{\lambda}. \qquad (*)$$

De este modo, *la esperanza matemática de la distribución exponencial es la magnitud inversa del parámetro* λ.

Hallemos la dispersión (cap. XII, § 1):

$$D(X)=\int_0^\infty x^2 f(x)\,dx-[M(X)]^2=\lambda\int_0^\infty x^2 e^{-\lambda x}\,dx-\frac{1}{\lambda^2}.$$

Integrando por partes, hallamos

$$\lambda\int_0^\infty x^2 e^{-\lambda x}\,dx=\frac{2}{\lambda^2}.$$

Por lo tanto,

$$D(X)=\frac{1}{\lambda^2}.$$

Hallemos la desviación cuadrática media, para lo cual extraemos la raíz cuadrada de la dispersión:

$$\sigma(X)=\frac{1}{\lambda}. \qquad (**)$$

Comparando (*) y (**), deducimos que

$$M(X)=\sigma(X)=\frac{1}{\lambda},$$

es decir, *la esperanza matemática y la desviación cuadrática media de la distribución exponencial son iguales entre sí.*

Ejemplo. La magnitud aleatoria continua está distribuida por la ley exponencial

$$f(x) = 5e^{-5x} \text{ para } x \geqslant 0; \quad f(x) = 0 \text{ para } x < 0.$$

Hallar la esperanza matemática, la desviación cuadrática media y la dispersión de X.

SOLUCION. Por los datos $\lambda = 5$. Por lo tanto,

$$M(X) = \sigma(X) = \frac{1}{\lambda} = \frac{1}{5} = 0{,}2;$$

$$D(X) = \frac{1}{\lambda^2} = \frac{1}{5^2} = 0{,}04.$$

Nota. 1. Supongamos que en la práctica se estudia una magnitud aleatoria distribuida exponencialmente, desconociéndose el parámetro λ. Si la esperanza matemática también es una incógnita, en tal caso se halla su estimación (valor aproximado), tomándose como tal el promedio muestral $\bar{x}$ (cap. XVI, § 5).

En este caso, el valor aproximado del parámetro λ se halla por la igualdad

$$\lambda^* = \frac{1}{\bar{x}}.$$

Nota 2. Supongamos que existen fundamentos para admitir que una magnitud aleatoria estudiada en la práctica tiene distribución exponencial. Para verificar esta hipótesis se halla la estimación de la esperanza matemática y la desviación cuadrática media, es decir, se hallan el promedio muestral y la desviación cuadrática media muestral (cap. XVI, §§ 5, 9). Puesto que la esperanza matemática y la desviación cuadrática media de la distribución exponencial son iguales entre sí, sus estimaciones deben diferenciarse muy poco. Si las estimaciones son próximas entre sí, los datos de observaciones corroboran la hipótesis de la distribución exponencial de la magnitud a estudiar; si las estimaciones se diferencian considerablemente, la hipótesis se rechaza.

La distribución exponencial se utiliza profusamente en las aplicaciones, en particular, en la teoría de la fiabilidad, uno de cuyos conceptos fundamentales es la función de fiabilidad.

§ 4. Función de fiabilidad

Denominemos *elemento* un dispositivo cualquiera, independientemente de que éste sea «simple», o «complejo».

Supongamos que el elemento comienza a trabajar en el instante $t_0 = 0$ y al final del tiempo de duración t ocurre el fallo.

Designemos por T una magnitud aleatoria continua, es decir, la duración del tiempo del trabajo sin fallo del elemento. Si el elemento trabajó sin fallo (hasta el fallo) un tiempo menor que t, tendremos, por lo tanto, que en el tiempo de duración t se produjo el fallo.

De este modo, la función integral

$$F(t) = P(T < t)$$

determina la *probabilidad de fallo* en el tiempo t. Por lo tanto, la probabilidad del trabajo sin fallo en ese tiempo, de duración t, es decir, la probabilidad del suceso opuesto $T > t$, es igual a

$$R(t) = P(T > t) = 1 - F(t). \qquad (*)$$

Se llama *función de fiabilidad* $R(t)$ la función que determina la probabilidad del trabajo sin fallo del elemento en el tiempo t:

$$R(t) = P(T > t).$$

§ 5. Ley exponencial de la fiabilidad

Generalmente la duración del tiempo del trabajo sin fallo del elemento tiene una distribución exponencial, cuya función integral es

$$F(t) = 1 - e^{-\lambda t}.$$

Por lo tanto, gracias a la correlación (*) del párrafo anterior, la función de fiabilidad, en el caso de la distrbución exponencial del tiempo del trabajo sin fallo del elemento, tiene la forma

$$R(t) = 1 - F(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

Se llama *ley exponencial de la fiabilidad* la función de fiabilidad, determinada por la igualdad

$$R(t) = e^{-\lambda t}, \qquad (*)$$

donde λ es la intensidad de fallos.

Como se deduce de la definición de función de fiabilidad (§ 4), esta fórmula permite hallar la probabilidad del trabajo sin fallo del elemento en un intervalo de tiempo de duración t, si el tiempo del trabajo sin fallo tiene distribución exponencial.

Ejemplo. El tiempo del trabajo sin fallo de un elemento está distribuido por una ley exponencial $f(t) = 0{,}02e^{-0{,}02t}$ para $t \geqslant 0$ (t, tiempo en horas). Hallar la probabilidad de que el elemento trabaje sin fallo 100 horas.

SOLUCION. Por los datos, la constante de la intensidad de fallos $\lambda = 0.02$. Utilizamos la fórmula (*):

$$R(100) = e^{-0{,}02 \cdot 100} = e^{-2} = 0{,}13534.$$

La probabilidad buscada de que el elemento trabaje sin fallo 100 horas, es aproximadamente igual a 0,14.

Nota. Si los fallos de los elementos en instantes fortuitos forman un flujo elemental, la probabilidad de que en el tiempo t no se produzca ningún fallo (cap. VI, § 6) es

$$P_t(0) = e^{-\lambda t},$$

lo que concuerda con la igualdad (*), puesto que λ en ambas fórmulas tiene el mismo sentido (constante de la intensidad de fallos).

§ 6. Propiedad característica de la ley exponencial de la fiabilidad

La ley exponencial de la fiabilidad es muy simple y conveniente para resolver problemas que surgen en la práctica. Muchas fórmulas de la teoría de la fiabilidad se simplifican considerablemente. Esto se debe a que la ley posee la siguiente propiedad importante: *la probabilidad del trabajo sin fallo de un elemento en el intervalo de tiempo, de duración t, no depende del tiempo de trabajo precedente hasta el origen del intervalo a considerar sino depende solamente de la duración del tiempo t* (para una intensidad de fallos λ dada).

Para demostrar la propiedad introducimos las denominaciones de los sucesos:

A — el trabajo sin fallo del elemento en el intervalo $(0, t_0)$ de duración t_0;

B — el trabajo sin fallo en el intervalo $(t_0, t_0 + t)$ de duración t.

En tal caso, AB es el trabajo sin fallo en el intervalo $(0, t_0 + t)$ de duración $t_0 + t$.

Hallamos las probabilidades de estos sucesos por la fórmula (*) (§ 5);

$$P(A) = e^{-\lambda t_0}, \quad P(B) = e^{-\lambda t},$$

$$P(AB) = e^{-\lambda(t_0+t)} = e^{-\lambda t_0} \cdot e^{-\lambda t}.$$

Hallamos la probabilidad condicional de que el elemento trabajará sin fallo en el intervalo $(t_0, t_0 + t)$ a condición de que ya haya trabajado sin fallo en el intervalo precedente $(0, t_0)$ (cap. III, § 5, nota 2):

$$P_A(B) = \frac{P(AB)}{P(A)} = \frac{e^{-\lambda t_0}e^{-\lambda t}}{e^{-\lambda t_0}} = e^{-\lambda t}.$$

Vemos que la fórmula obtenida no contiene t_0 y contiene sólo t. Esto significa, precisamente, que el tiempo de trabajo en el intervalo anterior no se manifiesta en la magnitud de la probabilidad del trabajo sin fallo en el intervalo siguiente, y depende sólo de la longitud del intervalo siguiente, lo que se quería demostrar.

El resultado obtenido se puede formular de un modo distinto. Comparando las probabilidades $P(B) = e^{-\lambda t}$ y $P_A(B) = e^{-\lambda t}$, se deduce que: la probabilidad condicional del trabajo sin fallo de un elemento en el intervalo de duración t, calculada suponiendo que el elemento trabajó sin fallo en el intervalo precedente, es igual a la probabilidad absoluta.

De este modo, en el caso de la ley exponencial de la fiabilidad, el trabajo sin fallo de un elemento «en el pasado» no se manifiesta en la magnitud de la probabilidad de su trabajo sin fallo «en un futuro próximo».

Nota. Se puede demostrar que *solamente* la distribución exponencial tiene la propiedad examinada. Por eso, si en la práctica la magnitud aleatoria a estudiar posee esta propiedad, ella está distribuida según una ley esponencial. Por ejemplo, al admitir que los meteoritos están distribuidos uniformemente en el espacio y en el tiempo, la probabilidad de impacto de un meteorito en una nave espacial no depende de que en la nave hayan hecho impacto o no meteoritos hasta el comienzo del intervalo de tiempo examinado. Por lo tanto, los instantes aleatorios de impacto de los meteoritos en la nave espacial están distribuidos por una ley exponencial.

Problemas

1. Escribir las funciones diferencial e integral de la distribución exponencial, si el parámetro $\lambda = 5$.

Respuesta $f(x) = 5e^{-5x}$ para $\geqslant 0$

$f(x) = 0$ para < 0;

$F(x) = 1 - e^{-5x}$.

2. La magnitud aleatoria continua X está distribuida por una ley exponencial: $f(x) = 5e^{-5x}$ para $x \geqslant 0$, $f(x) = 0$ para $x < 0$. Hallar la probabilidad de que como resultado del experimento, X cae en el intervalo (0,4; 1)

Respuesta $P(0{,}4 < X < 1) = 0{,}13$.

3. La magnitud aleatoria continua X está distribuida por una ley exponencial $f(x) = 4e^{-4x}$ $(x > 0)$. Hallar la esperanza matemática, la desviación cuadrática media y la dispersión de X.

Respuesta $M(X) = \sigma(X) = 0{,}25$;
$D(X) = 0{,}0625$.

4. El tiempo del trabajo sin fallo de un elemento está distribuido por la ley exponencial $f(t) = 0{,}01 \cdot e^{-0{,}01t}$ $(t > 0)$, donde t es el tiempo en horas. Hallar la probabilidad de que el elemento trabaje sin fallo 100 horas.

Respuesta $R(100) = 0{,}37$.

Capítulo catorce

SISTEMA DE DOS MAGNITUDES ALEATORIAS

§ 1. Noción de sistema de varias magnitudes aleatorias

Hasta ahora consideramos las magnitudes aleatorias, cuyos valores posibles se determinaron por un número. Estas magnitudes se llaman unidimensionales. Por ejemplo, el número de puntos, que puede caer al tirar un dado, es una magnitud unidimensional discreta; la distancia desde un cañón hasta el lugar de impacto del proyectil es una magnitud aleatoria unidimensional continua.

Además de las magnitudes aleatorias unidimensionales, se estudian las magnitudes, cuyos valores posibles se determinan por dos, tres, . . ., n números. Estas magnitudes se llaman respectivamente bidimensionales, tridimensionales, . . ., n -dimensionales.

Designaremos por (X, Y) la magnitud aleatoria bidimensional. Cada una de las magnitudes X e Y se llama componente; ambas magnitudes X e Y, consideradas simultáneamente, forman un sistema de dos magnitudes aleatorias. Análogamente, la magnitud n-dimensional se puede considerar como un sistema de n magnitudes aleatorias. Por

ejemplo, la magnitud tridimensional (X, Y, Z) determina un sistema de tres magnitudes aleatorias X, Y y Z.

Ejemplo. Una máquina automática estampa planchas de acero. Si las dimensiones a controlar son la longitud X y la anchura Y, tendremos una magnitud aleatoria bidimensional (X, Y); si se controla también la altura Z, tendremos una magnitud tridimensional (X, Y, Z).

Geométricamente, una magnitud aleatoria bidimensional (X, Y) se puede interpretar como un punto fortuito $M (X, Y)$ en el plano (es decir, como un punto de coordenadas fortuitas), o bien como un vector $\overline{OM}$. La magnitud aleatoria tridimensional puede interpretarse geométricamente como un punto $M (X, Y, Z)$ en el espacio tridimensional, o como un vector $\overline{OM}$.

Conviene distinguir las magnitudes aleatorias polidimensionales discretas (las componentes de estas magnitudes son discretas) y las continuas (las componentes de estas magnitudes son continuas).

§ 2. Ley de distribución de las probabilidades de una magnitud aleatoria bidimensional discreta

Se llama *ley de distribución* de una magnitud aleatoria bidimensional discreta la enumeración de los valores posibles de esta magnitud (es decir, del par de números (x_i, y_j) y de sus probabilidades $p(x_i, y_j)$ $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$. De ordinario, la ley de distribución se da en forma de una tabla con doble entrada (tabla 2).

La primera línea de la tabla contiene todos los valores posibles de la componente X y la primera columna, todos los valores posibles de la componente Y. En la casilla situada en la intersección de la «columna x_i» y la «línea y_j», se indica la probabilidad $p(x_i, y_j)$ de que la magnitud aleatoria bidimensional toma el valor (x_i, y_j).

Puesto que los sucesos $(X = x_i, Y = y_j)$, $(i = 1, 2, \ldots \ldots, n;\ j = 1, 2, \ldots, m)$ forman un grupo completo (cap. II, § 2), la suma de las probabilidades, alojadas en todas las casillas de la tabla, es igual a la unidad.

Conociendo la ley de distribución de una magnitud aleatoria bidimensional discreta se pueden hallar las leyes de distribución de cada una de las componentes. En efecto, por ejemplo, los sucesos $(X = x_1;\ Y = y_1)$, $(X = x_1;\ Y = y_2)$, $\ldots$, $(X = x_1;\ Y = y_m)$ son mutuamente excluyen-

Tabla 2

$Y \backslash X$	x_1	x_2	...	x_i	...	x_n
y_1	$p(x_1, y_1)$	$p(x_1, y_1)$		$p(x_i, y_1)$		$p(x_n, y_1)$
...	...	...	...	...	...	...
y_j	$p(x_1, y_j)$	$p(x_2, y_j)$	...	$p(x_i, y_j)$	...	$p(x_n, y_j)$
...	...	...	...	...	...	...
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	...	$p(x_i, y_m)$	...	$p(x_n, y_m)$

tes, por eso la probabilidad $P(x_1)$ de que X tome el valor x_1, por el teorema de la adición, es:

$$P(x_1) = p(x_1, y_1) + p(x_1, y_2) + \ldots + p(x_1, y_m).$$

Por consiguiente, la probabilidad de que X tome el valor x_1, es igual a la suma de las probabilidades de la «columna x_1». En el caso general, para hallar la probabilidad $P(X = x_i)$ hay que sumar las probabilidades de la columna x_i. Análogamente, sumando las probabilidades de la «línea y_j», obtenemos la probabilidad $P(Y = y_j)$.

Ejemplo. Hallar las leyes de distribución de las componentes de una magnitud aleatoria bidimensional, prefijada por la ley de distribución (tabla 3).

Tabla 3

$Y \backslash X$	x_1	x_2	x_3
y_1	0,10	0,30	0,20
y_2	0,06	0,18	0,16

SOLUCION. Sumando las probabilidades por columnas, obtenemos las probabilidades de los valores posibles de X: $p(x_1) = 0{,}16$; $p(x_2) = 0{,}48$; $p(x_3) = 0{,}36$. Escribimos la ley de distribución de la componente X:

X	x_1	x_2	x_3.
p	0,16	0,48	0,36

Verificación: $0{,}16 + 0{,}48 + 0{,}36 = 1$.

Sumando las probabilidades por líneas, obtenemos las probabilidades de los valores posibles de Y: $p(y_1) = 0{,}60$; $p(y_2) = 0{,}40$. Escribimos la ley de distribución de la componente Y:

Y	y_1	y_2
p	0,60	0,40.

Verificación: $0{,}60 + 0{,}40 = 1$.

§ 3. Función integral de distribución de una magnitud aleatoria bidimensional

Examinemos una magnitud aleatoria bidimensional (X, Y) (indistintamente discreta o continua). Supongamos que x, y es un par de números reales. La probabilidad del suceso consistente en que X tome un valor menor que x, y al mismo tiempo Y tome un valor menor que y, lo designamos por $F(x, y)$. Si x e y varían, en general, variará también $F(x, y)$, es decir, $F(x, y)$ es una función de x e y.

Se llama *función integral de distribución* de una magnitud aleatoria bidimensional (X, Y) la función $F(x, y)$ que determina para cada par de números x, y la probabilidad de que X tome un valor menor que x y, en este caso, Y tome un valor menor que y:

$$F(x, y) = P(X < x,\ Y < y).$$

Geométricamente, esta igualdad se puede interpretar así: $F(x, y)$ es la probabilidad de que el punto fortuito (X, Y) caiga en un cuadrante infinito de vértice (x, y), ubicado a la izquierda y debajo de este vértice (fig. 13).

Ejemplo. Hallar la probabilidad de que debido al experimento la componente X de una magnitud aleatoria bidimensional (X, Y), tome un valor $X < 2$ y en este caso la

componente Y tome un valor $Y < 3$, si se conoce la función integral del sistema

$$F(x, y) = \left(\frac{1}{\pi} \operatorname{arctg} \frac{x}{2} + \frac{1}{2}\right) \cdot \left(\frac{1}{\pi} \operatorname{arctg} \frac{y}{3} + \frac{1}{2}\right).$$

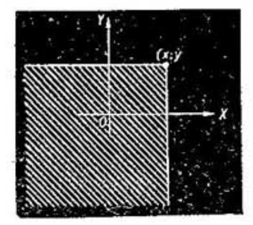


Fig. 13.

SOLUCION. Por definición de la función integral de una magnitud aleatoria bidimensional

$$F(x, y) = P(X < x, Y < y).$$

Poniendo $x = 2$, $y = 3$, obtenemos la probabilidad buscada

$$P(X < 2, Y < 3) = F(2, 3) = \left(\frac{1}{\pi} \operatorname{arctg} \frac{2}{2} + \frac{1}{2}\right) \times$$
$$\times \left(\frac{1}{\pi} \operatorname{arctg} \frac{3}{3} + \frac{1}{2}\right) = \left(\frac{1}{\pi} \cdot \frac{\pi}{4} + \frac{1}{2}\right) \times$$
$$\times \left(\frac{1}{\pi} \cdot \frac{\pi}{4} + \frac{1}{2}\right) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}.$$

§ 4. Propiedades de la función integral de una magnitud aleatoria bidimensional

Propiedad 1. *Los valores de la función integral satisfacen la doble desigualdad*

$$0 \leqslant F(x, y) \leqslant 1.$$

DEMOSTRACION. La propiedad resulta de la definición de función integral como una probabilidad: la proba-

bilidad siempre es un número no negativo, no mayor que la unidad.

Propiedad 2. *$F(x, y)$ es una función no decreciente por cada argumento, o sea,*

$$F(x_2, y) \geqslant F(x_1, y), \quad \text{si} \quad x_2 > x_1;$$
$$F(x, y_2) \geqslant F(x, y_1), \quad \text{si} \quad y_2 > y_1.$$

DEMOSTRACION. Vamos a demostrar que $F(x, y)$ es una función no decreciente por el argumento x. El suceso, consistente en que la componente X tome un valor menor que x_2 y al mismo tiempo la componente $Y < y$, se puede dividir en dos sucesos mutuamente excluyentes:

1) X toma un valor menor que x_1 y, en ese caso, $Y < y$ con probabilidad $P(X < x_1, Y < y)$;

2) X toma un valor que satisface la desigualdad $x_1 \leqslant X < x_2$ y con eso $Y < y$ con probabilidad $P(x_1 \leqslant X < x_2, Y < y)$.

Según el teorema de la adición tenemos

$$P(X < x_2, Y < y) = P(X < x_1, Y < y) + P(x_1 \leqslant X < x_2, Y < y).$$

De donde

$$P(X < x_2, Y < y) - P(X < x_1, Y < y) = P(x_1 \leqslant X < x_2, Y < y),$$

o bien

$$F(x_2, y) - F(x_1, y) = P(x_1 \leqslant X < x_2, Y < y)$$

Puesto que toda probabilidad es un número no negativo, entonces

$$F(x_2, y) - F(x_1, y) \geqslant 0,$$

o bien

$$F(x_2, y) \geqslant F(x_1, y),$$

lo que se quería demostrar.

La propiedad se hace evidente, si se interpreta geométricamente la función integral como la probabilidad de que un punto aleatorio caiga en un cuadrante infinito de vértice (x, y) (fig. 13). Al crecer x el límite de derecha de este cuadrante se mueve hacia la derecha; en este caso, la probabili-

dad de que el punto aleatorio caiga en el «nuevo» cuadrante, evidentemente, no puede disminuir.

Análogamente se demuestra que $F(x, y)$ es una función no decreciente por el argumento y.

Propiedad 3. *Se producen las correlaciones límites*

1) $F(-\infty, y) = 0$,
2) $F(x, -\infty) = 0$,
3) $F(-\infty, -\infty) = 0$,
4) $F(\infty, \infty) = 1$.

DEMOSTRACION. 1) $F(-\infty, y)$ es la probabilidad del suceso $X < -\infty$ e $Y < y$; pero tal suceso es imposible (puesto que el suceso $X < -\infty$ es imposible), por lo tanto, la probabilidad de este suceso es igual a cero.

La propiedad se hace evidente, si se recurre a la interpretación geométrica: cuando $x \to -\infty$ el límite de derecha del cuadrante infinito (fig. 13) se desplaza ilimitadamente hacia la izquierda y, en este caso, la probabilidad de que el punto aleatorio caiga en el cuadrante tiende a cero.

2) El suceso $Y < -\infty$ es imposible, por eso $F(x, -\infty) = 0$.

3) El suceso $X < -\infty$ e $Y < -\infty$ es imposible, por eso $F(-\infty, -\infty) = 0$.

4) El suceso $X < \infty$ e $Y < \infty$ es cierto, por lo tanto, la probabilidad de este suceso es $F(\infty, \infty) = 1$.

La propiedad se hace evidente, si se tiene en cuenta que, para $x \to \infty$ e $y \to \infty$ el cuadrante infinito (fig. 13) se convierte en todo el plano XOY y, por lo tanto, la caída del punto aleatorio (X, Y) en este plano, como resultado del experimento, es un suceso cierto.

Propiedad 4. a) *Cuando $y = \infty$ la función integral del sistema deviene una función integral de la componente X:*

$$F(x, \infty) = F_1(x).$$

b) *Cuando $x = \infty$ la función integral del sistema deviene una función integral de la componente Y:*

$$F(\infty, y) = F_2(y).$$

DEMOSTRACION. a) Dado que el suceso $Y < \infty$ es cierto, tendremos que $F(x, \infty)$ define la probabilidad del suceso $X < x$, es decir, es la función integral de la componente X

b) Se demuestra de manera análoga.

§ 5. Probabilidad de que un punto aleatorio caiga en una semizona

Utilizando la función integral del sistema de magnitudes aleatorias X e Y, se halla fácilmente la probabilidad de que como resultado del experimento el punto aleatorio caiga en la semizona $x_1 < X < x_2$ e $Y < y$ (fig. 14, *a*), o en la semizona $X < x$ e $y_1 < Y < y_2$ (fig. 14, *b*).

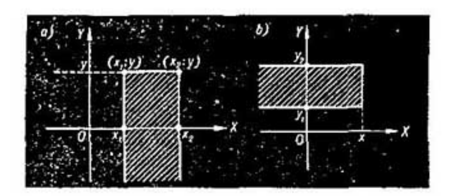


Fig. 14.

Restando de la probabilidad de que el punto aleatorio caiga en el cuadrante de vértice (x_2, y) la probabilidad de que el punto caiga en cuadrante de vértice (x_1, y) (fig. 14, *a*) obtenemos

$$P(x_1 < X < x_2,\ Y < y) = F(x_2, y) - F(x_1, y).$$

Análogamente tenemos

$$P(X < x,\ y_1 < Y < y_2) = F(x, y_2) - F(x, y_1).$$

Por consiguiente, la probabilidad de que el punto aleatorio caiga en una semizona igual al incremento de la función integral según cada uno de los agrumentos.

§ 6. Probabilidad de que un punto aleatorio caiga en un rectángulo

Examinemos el rectángulo $ABCD$ con los lados paralelos a los ejes de coordenadas (fig. 15). Supongamos que las ecuaciones de los lados son:

$$X = x_1, \quad X = x_2, \quad Y = y_1 \text{ e } Y = y_2.$$

Hallemos la probabilidad de que el punto aleatorio (X, Y) caiga en este rectángulo. La probabilidad buscada se puede hallar, por ejemplo, así: de la probabilidad de que el punto aleatorio caiga en la semizona AB de rayado

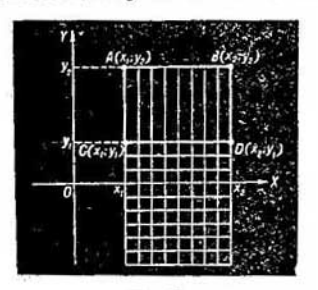


Fig. 15.

vertical (esta probabilidad es igual a $F(x_2, y_2) - F(x_1, y_2)$ se resta la probabilidad de que el punto caiga en la semizona CD de rayado horizontal (esta probabilidad es igual a $F(x_2, y_1) - F(x_1, y_1)$:

$$P(x_1 < X < x_2, y_1 < Y < y_2) = [F(x_2, y_2) - F(x_1, y_2)] - [F(x_2, y_1) - F(x_1, y_1)]. \quad (*)$$

Ejemplo. Hallar la probabilidad de que el punto aleatorio (X, Y) caiga en el rectángulo, limitado por las rectas $x = \frac{\pi}{6}$, $x = \frac{\pi}{2}$, $y = \frac{\pi}{4}$, $y = \frac{\pi}{3}$, si se conoce la función integral

$$F(x, y) = \operatorname{sen} x \cdot \operatorname{sen} y \left(0 \leqslant x \leqslant \frac{\pi}{2},\ 0 \leqslant y \leqslant \frac{\pi}{2}\right).$$

SOLUCION. Poniendo $x_1 = \frac{\pi}{6}$, $x_2 = \frac{\pi}{2}$, $y_1 = \frac{\pi}{2}$, $y = \frac{\pi}{3}$ en la fórmula (*), obtenemos:

$$P\left(\frac{\pi}{6} < X < \frac{\pi}{2}, \frac{\pi}{4} < Y < \frac{\pi}{3}\right) = \left[F\left(\frac{\pi}{2}, \frac{\pi}{3}\right) - F\left(\frac{\pi}{6}, \frac{\pi}{3}\right)\right] - \left[F\left(\frac{\pi}{2}, \frac{\pi}{4}\right) - F\left(\frac{\pi}{6}, \frac{\pi}{4}\right)\right] =$$

$$=\left[\operatorname{sen}\frac{\pi}{2}\cdot\operatorname{sen}\frac{\pi}{3}-\operatorname{sen}\frac{\pi}{6}\cdot\operatorname{sen}\frac{\pi}{3}\right]-$$
$$-\left[\operatorname{sen}\frac{\pi}{2}\cdot\operatorname{sen}\frac{\pi}{4}-\operatorname{sen}\frac{\pi}{6}\cdot\operatorname{sen}\frac{\pi}{4}\right]=$$
$$=\left[\frac{\sqrt{3}}{2}-\frac{1}{2}\cdot\frac{\sqrt{3}}{2}\right]-\left[\frac{\sqrt{2}}{2}-\frac{1}{2}\cdot\frac{\sqrt{2}}{2}\right]=$$
$$=\frac{\sqrt{3}-\sqrt{2}}{4}=0{,}08.$$

§ 7. Función diferencial de una magnitud aleatoria bidimensional continua (densidad de probabilidad bidimensional)

Hemos fijado una magnitud aleatoria bidimensional mediante una función integral. La magnitud aleatoria bidimensional continua también se puede fijar utilizando la función diferencial de distribución. Aquí y en adelante vamos a suponer que la función integral es continua en todas partes y tiene en todas partes (con excepción, puede ser, de un número finito de curvas) una derivada parcial mixta continua de segundo orden.

Se llama *función diferencial de distribución* $f(x, y)$ de la magnitud aleatoria bidimensional continua (X, Y) la derivada parcial mixta segunda de la función integral:

$$f(x, y)=\frac{\partial^2 F(x, y)}{\partial x\, \partial y}.$$

Esta función se puede interpretar geométricamente como una superficie que se denomina *superficie de distribución*.

Ejemplo. Hallar la función diferencial $f(x, y)$ del sistema de magnitudes aleatorias (X, Y) según la función integral conocida

$$F(x, y)=\operatorname{sen} x\cdot\operatorname{sen} y\left(0\leqslant x\leqslant\frac{\pi}{2},\ 0\leqslant y\leqslant\frac{\pi}{2}\right).$$

SOLUCION. Por definición de la función diferencial de un sistema de magnitudes aleatorias

$$f(x, y)=\frac{\partial^2 F}{\partial x\, \partial y}.$$

Hallamos la derivada parcial por x de la función integral

$$\frac{\partial F}{\partial x}=\cos x\cdot\operatorname{sen} y.$$

Del resultado obtenido hallamos la derivada parcial respecto de y, debido a lo cual obtenemos la función diferencial buscada

$$f(x, y) = \frac{\partial^2 F}{\partial x\, \partial y} = \cos x \cdot \cos y \left(0 \leqslant x \leqslant \frac{\pi}{2},\ 0 < y \leqslant \frac{\pi}{2}\right).$$

§ 8. Hallazgo de la función integral de distribución por la función diferencial conocida

Conociendo la función diferencial $f(x, y)$ se puede hallar la función integral $F(x, y)$ por la fórmula

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)\, dx\, dy,$$

lo que resulta directamente de la definición de función diferencial.

Ejemplo. Hallar la función integral de distribución de una magnitud aleatoria bidimensional por la función diferencial dada $f(x, y) = \frac{1}{\pi^2 (1+x^2)(1+y^2)}$.

SOLUCION. Utilizamos la fórmula

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)\, dx\, dy.$$

Poniendo aquí $f(x, y) = \frac{1}{\pi^2 (1+x^2)(1+y^2)}$, obtenemos

$$F(x, y) = \frac{1}{\pi^2} \int_{-\infty}^{y} \left(\frac{1}{1+y^2} \int_{-\infty}^{x} \frac{1}{1+x^2} \right) dy =$$

$$= \frac{1}{\pi^2} \int_{-\infty}^{y} \frac{1}{1+y^2} \left(\operatorname{arctg} x + \frac{\pi}{2} \right) dy =$$

$$= \left(\frac{1}{\pi} \operatorname{arctg} x + \frac{1}{2} \right) \frac{1}{\pi} \int_{-\infty}^{y} \frac{dy}{1+y^2} =$$

$$= \left(\frac{1}{\pi} \operatorname{arctg} x + \frac{1}{2} \right) \left(\frac{1}{\pi} \operatorname{arctg} y + \frac{1}{2} \right).$$

§ 9. Sentido probabilístico de la función diferencial de una magnitud aleatoria bidimensional

La probabilidad de que el punto aleatorio (X, Y) caiga en el rectángulo $ABCD$ (fig. 16) es igual a (§ 6)

$$P(x_1 < X < x_2,\ y_1 < Y < y_2) = \\ = [F(x_2, y_2) - F(x_1, y_2)] - \\ - [F(x_2, y_1) - F(x_1, y_1)].$$

Para abreviar designamos el primer miembro de la igualdad por P_{ABCD} y aplicando al segundo miembro el teorema

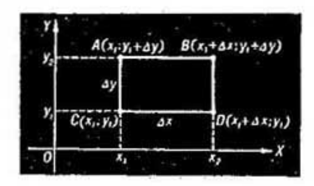


Fig. 16.

de Lagrange, obtenemos

$$P_{ABCD} = F''_{xy}(\xi, \eta) \cdot \Delta x \cdot \Delta y,$$

donde

$$x_1 < \xi < x_2 \quad \Delta x = x_2 - x_1,$$
$$y_1 < \eta < y_2, \quad \Delta y = y_2 - y_1.$$

De donde

$$F''_{xy}(\xi, \eta) = \frac{P_{ABCD}}{\Delta x \cdot \Delta y}. \qquad (*)$$

o bien

$$f(\xi, \eta) = \frac{P_{ABCD}}{\Delta x \cdot \Delta y}. \qquad (**)$$

Tomando en consideración que el producto $\Delta x \cdot \Delta y$ es igual a la superficie del rectángulo $ABCD$, deducimos que: $f(\xi, \eta)$ es la relación de la probabilidad de que el punto aleatorio caiga en el rectángulo $ABCD$ a la superficie de este rectángulo.

Pasamos ahora en la igualdad (**) al límite para $\Delta x \to 0$ y $\Delta y \to 0$. En tal caso, $\xi \to x$, $\eta \to y$ y, por lo tanto, $f(\xi, \eta) \to f(x, y)$.

De este modo, la función $f(x, y)$ se puede considerar como límite de la relación de la probabilidad de que el punto aleatorio caiga en el rectángulo (de lados Δx y Δy) a la superficie de este rectángulo, cuando ambos lados del rectángulo tienden a cero.

§ 10. Probabilidad de que un punto aleatorio caiga en una región arbitraria

Escribamos la correlación (**) del § 9 así:

$$f(\xi, \eta) \cdot \Delta x \cdot \Delta y = P_{ABCD}.$$

De donde se deduce: el producto $f(\xi, \eta) \cdot \Delta x \cdot \Delta y$ es la probabilidad de que un punto aleatorio caiga en el rectángulo de lados Δx y Δy.

Supongamos que en plano XOY se ha dado una región arbitraria D. Designemos el suceso consistente en la incidencia del punto aleatorio en esta región por: $(X, Y) \subset D$.

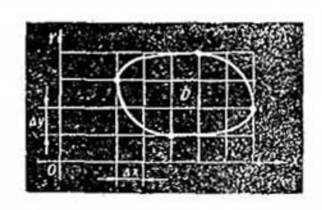


Fig. 17.

Descompongamos la región D en n regiones elementales por rectas paralelas al eje OY que se encuentran a la distancia Δx una de otra y rectas parelales al eje OX distanciados entre sí en Δy (fig. 17) (para simpleza se supone que estas rectas intersectan el contorno de la región no más que en dos puntos).

Puesto que los sucesos consistentes en la incidencia del punto aleatorio en regiones elementales, son mutuamente excluyentes, la probabilidad de incidencia en la región D

es aproximadamente (¡la suma de las regiones elementales es aproximadamente igual a la región D!) igual a la suma de las probabilidades de que el punto caiga en las regiones elementales:

$$P((X, Y) \subset D) \approx \sum_{i=1}^{n} f(\xi_i, \eta_i) \cdot \Delta x \cdot \Delta y.$$

Pasando al límite para $\Delta x \to 0$ y $\Delta y \to 0$, obtenemos

$$P((X, Y) \subset D) = \iint_{(D)} f(x, y)\, dx\, dy. \qquad (*)$$

Así pues, para calcular la probabilidad de que el punto aleatorio (X, Y) caiga en la región D, es suficiente hallar la integral doble de la función diferencial en el campo D.

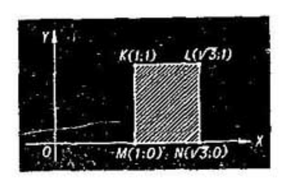


Fig. 18.

Geométricamente la igualdad (*) se puede interpretar así: la probabilidad de que el punto aleatorio (X, Y) caiga en la región D es igual al volumen del cuerpo, limitado en la parte superior por la superficie $Z = f(x, y)$, siendo su base la proyección de esta superficie sobre el plano XOY.

Nota. La expresión subintegral $f(x, y)\, dx\, dy$ se llama *elemento de probabilidad.* Como se deduce de lo anterior, el elemento de probabilidad determina la probabilidad de que un punto aleatorio caiga en un rectángulo elemental de lados dx y dy.

Ejemplo. Se ha dado la función diferencial de una magnitud aleatoria bidimensional

$$f(x, y) = \frac{1}{\pi^2 (1 + x^2)(1 + y^2)}.$$

Hallar la probabilidad de que un punto aleatorio caiga en el rectángulo (fig. 18) de vértices $K(1; 1)$, $L(\sqrt{3}; 1)$, $M(1; 0)$ y $N(\sqrt{3}; 0)$.

SOLUCION. La probabilidad buscada es

$$P((X, Y) \subset D) = \iint_{(D)} \frac{1}{\pi^2 (1+x^2)(1+y^2)} \, dx\, dy =$$

$$= \frac{1}{\pi^2} \int_0^1 \left[\frac{1}{1+y^2} \int_1^{\sqrt{3}} \frac{dx}{1+x^2} \right] dy = \frac{1}{\pi^2} \times$$

$$\times \operatorname{arctg} x \Big|_1^{\sqrt{3}} \int_0^1 \frac{dy}{1+y^2} = \frac{1}{\pi^2} \left(\frac{\pi}{3} - \frac{\pi}{4} \right) \times$$

$$\times \operatorname{arctg} y \Big|_0^1 = \frac{1}{\pi^2} \cdot \frac{\pi}{12} \cdot \frac{\pi}{4} = \frac{1}{48}.$$

§ 11. Propiedades de la función diferencial de una magnitud aleatoria bidimensional

Propiedad 1. *La función diferencial no es negativa*

$$f(x, y) \geqslant 0.$$

DEMOSTRACION. La probabilidad de que un punto aleatorio caiga en un rectángulo de lados Δx y Δy es un número no negativo; el área de este rectángulo es un número positivo. Por lo tanto, la relación de estos dos números, y, en consecuencia, también sus límites (para $\Delta x \to 0$ y $\Delta y \to \to 0$) que es igual a $f(x, y)$ (§ 9), es un número no negativo, es decir

$$f(x, y) \geqslant 0.$$

Cabe hacer notar que la propiedad se deduce directamente del hecho de que $F(x, y)$ es una función no decreciente de sus argumentos (§ 4).

Propiedad 2. *La integral impropia doble de límites infinitos de una función diferencial es igual a la unidad:*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

DEMOSTRACION. Los límites infinitos de integración indican que el campo de integración es todo el plano xOy;

puesto que el suceso consistente en que un punto aleatorio caiga durante el experimento en el plano xOy es cierto, la probabilidad de este suceso (que precisamente se determina por la integral impropia doble de la función diferencial) es igual a la unidad, es decir,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$

§ 12. Hallazgo de las funciones diferenciales de las componentes de una magnitud aleatoria bidimensional

Supongamos que se conoce la función diferencial de un sistema de dos magnitudes aleatorias. Tratemos de hallar las funciones diferenciales de cada una de las componentes.

Al principio hallamos la función diferencial $f_1(x)$ de la componente X. Designemos por $F_1(x)$ la función integral de la componente X. Por definición de la función diferencial de una magnitud aleatoria unidimensional

$$f_1(x) = \frac{dF_1(x)}{dx}.$$

Teniendo en cuenta las correlaciones

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x, y)\, dx\, dy \qquad (\S\ 8)$$

$$F_1(x) = F(x, \infty) \qquad (\S\ 4),$$

hallamos

$$F_1(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy.$$

Diferenciando ambos miembros de esta igualdad respecto de x, obtenemos

$$\frac{dF_1}{dx} = \int_{-\infty}^{\infty} f(x, y)\, dy,$$

o bien

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy. \qquad (^*)$$

Análogamente se halla la función diferencial de la componente Y:

$$f_2(y)=\int_{-\infty}^{\infty} f(x,\ y)\,dx. \qquad (^{**})$$

De este modo, *la función diferencial de una de las componentes es igual a la integral impropia de límites infinitos de la función diferencial del sistema; además, la variable de integración corresponde a la otra componente.*

Ejemplo. La magnitud aleatoria bidimensional (X, Y) está prefijada por la función diferencial

$$f(x,\ y)=\begin{cases} \frac{1}{6\pi} & \text{para } \frac{x^2}{9}+\frac{y^2}{4}<1, \\ 0 & \text{para } \frac{x^2}{9}+\frac{y^2}{4}>1. \end{cases}$$

Hallar las funciones diferenciales de las componentes X e Y.

SOLUCION. Hallamos la función diferencial de la componente X por la fórmula (*)

$$f_1(x)=\int_{-\infty}^{\infty}\frac{1}{6\pi}\,dy=\frac{1}{6\pi}\int_{-2\sqrt{1-\frac{x^2}{9}}}^{2\sqrt{1-\frac{x^2}{9}}} dy=\frac{2}{6\pi}\times$$

$$\times\int_{0}^{2\sqrt{1-\frac{x^2}{9}}} dy=\frac{2}{9\pi}\sqrt{9-x^2}\ .$$

Por consiguiente,

$$f_1(x)=\begin{cases} \frac{2}{9\pi}\sqrt{9-x^2} & \text{para } |x|<3, \\ 0 & \text{para } |x|\geqslant 3. \end{cases}$$

Análogamente, utilizando la fórmula (**), hallamos la función diferencial de la componente Y:

$$f_2(y)=\begin{cases} \frac{1}{2\pi}\sqrt{4-y^2} & \text{para } |y|<2, \\ 0 & \text{para } |y|\geqslant 2. \end{cases}$$

Recomendamos al lector, para control, cerciorarse personalmente de que las funciones halladas satisfacen las

correlaciones

$$\int_{-\infty}^{\infty} f_1(x)\,dx = 1 \quad \text{y} \quad \int_{-\infty}^{\infty} f_2(y)\,dy = 1.$$

§ 13. Leyes condicionales de distribución de las componentes de un sistema de magnitudes aleatorias discretas

Hemos establecido que si los sucesos A y B son dependientes, la probabilidad condicional del suceso B se diferencia de su probabilidad absoluta. En este caso (cap. III, § 5, nota 2)

$$P_A(B) = \frac{P(AB)}{P(A)}. \qquad (*)$$

También para las magnitudes aleatorias se produce una situación análoga. Para caracterizar la dependencia entre las componentes de una magnitud aleatoria bidimensional, introducimos el concepto de distribución condicional.

Consideremos la magnitud aleatoria bidimensional discreta (X, Y). Supongamos que los valores posibles de las componentes sean

$$x_1, x_2, \ldots, x_n; \quad y_1, y_2, \ldots, y_m.$$

Admitamos que como resultado de la prueba la magnitud Y ha tomado el valor $Y = y_1$; en ese caso, X toma uno de sus valores posibles x_1 o $x_2, \ldots,$ o bien x_n. Designemos la probabilidad condicional de que X toma, por ejemplo, el valor x_1 a condición de que $Y = y_1$, por $p(x_1 \mid y_1)$. En general, esta probabilidad no será igual a la probabilidad absoluta $p(x_1)$.

En el caso general, las probabilidades condicionales de las componentes las designaremos así:

$$p(x_i \mid y_j) \quad (i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, m).$$

Se llama *distribución condicional* de la componente X para $Y = y_j$ el conjunto de probabilidades condicionales

$$p(x_1 \mid y_j),\ p(x_2 \mid y_j), \ldots, p(x_n \mid y_j),$$

calculadas suponiendo que el suceso $Y = y_j$ (j tiene el mismo valor para todos los valores de X) ya ocurrió.

Análogamente se determina la distribución condicional de la componente Y.

Conociendo la ley de distribución de una magnitud aleatoria bidimensional discreta, se pueden mediante la fórmula (*) calcular las leyes convencionales de distribución de las componentes. Por ejemplo, la ley condicional de distribucion de X, suponiendo que el suceso $Y = y_1$ ya ocurrió, puede ser hallada por la fórmula

$$p(x_i \mid y_1) = \frac{p(x_i,\ y_1)}{p(y_1)},$$

$$(i = 1, 2, \ldots, n)$$

En el caso general, las leyes condicionales de distribución de la componente X se determinan por la correlación

$$p(x_i \mid y_j) = \frac{p(x_i,\ y_j)}{p(y_j)}. \qquad (^{**})$$

De manera análoga se hallan las leyes condicionales de distribución de la componente Y:

$$p(y_j \mid x_i) = \frac{p(x_i,\ y_j)}{p(x_i)}. \qquad (^{***})$$

Nota. La suma de las probabilidades de una distribución condicional es igual a la unidad. En efecto, puesto que para y_j fijado tenemos (§ 2) $\sum_{i=1}^{n} p(x_i,\ y_j) = p(y_j)$, entonces

$$\sum_{i=1}^{n} p(x_i \mid y_j) = \sum_{i=1}^{n} \frac{p(x_i, y_j)}{p(y_j)} = \frac{p(y_j)}{p(y_j)} = 1.$$

Análogamente se demuestra que, para x_i fijado,

$$\sum_{j=1}^{m} p(y_j \mid x_i) = 1.$$

Esta propiedad de las distribuciones condicionales se utiliza para verificar los cálculos.

Ejemplo. Una magnitud aleatoria bidimensional discreta se prefija por la tabla 4.

Hallar la ley condicional de distribución de la componente X a condición de que la componente Y tome el valor y_1.

Tabla 4

$Y \backslash X$	x_1	x_2	x_3
y_1	0,10	0,30	0,20
y_2	0,06	0,18	0,16

SOLUCION. La ley buscada se determina por el conjunto de probabilidades condicionales siguientes:

$$p(x_1 \mid y_1), \; p(x_2 \mid y_1), \; p(x_3 \mid y_1).$$

Utilizando la fórmula (*) y tomando en consideración que $p(y_1) = 0{,}60$ (pág. 171), tenemos:

$$p(x_1 \mid y_1) = \frac{p(x_1, y_1)}{p(y_1)} = \frac{0{,}10}{0{,}60} = \frac{1}{6};$$

$$p(x_2 \mid y_1) = \frac{p(x_2, y_1)}{p(y_1)} = \frac{0{,}30}{0{,}60} = \frac{1}{2};$$

$$p(x_3 \mid y_1) = \frac{p(x_3, y_1)}{p(y_1)} = \frac{0{,}20}{0{,}60} = \frac{1}{3}.$$

Sumando las probabilidades condicionales halladas, nos cercioramos de que la suma obtenida es igual a la unidad, como debe ser (de acuerdo con la nota de la pág. 186):

$$\frac{1}{6} + \frac{1}{2} + \frac{1}{3} = 1.$$

§ 14. Leyes condicionales de distribución de las componentes de un sistema de magnitudes aleatorias continuas

Supongamos que (X, Y) es una magnitud aleatoria bidimensional continua. Se llama *función diferencial condicional* $\varphi(x \mid y)$ de la componente X, para un valor dado de $Y = y$, la relación entre la función diferencial $f(x, y)$ de un sistema y la función diferencial $f_2(y)$ de la componente Y:

$$\varphi(x \mid y) = \frac{f(x, y)}{f_2(y)}. \qquad (*)$$

Cabe señalar que la función condicional $\varphi(x \mid y)$ se distingue de la función diferencial absoluta $f_1(x)$ en que $\varphi(x \mid y)$ da una distribución de X a condición de que la componente Y ha tomado un valor $Y = y$. La función $f_1(x)$ da una distribución de X independientemente de cuáles de los valores posibles ha tomado la componente Y.

Análogamente se determina la función diferencial condicional de la componente Y para un valor dado de $X = x$:

$$\psi(y \mid x) = \frac{f(x, y)}{f_1(x)}. \qquad (**)$$

Si se conoce la función diferencial $f(x, y)$ de un sistema, las funciones diferenciales condicionales de las componentes pueden ser halladas en virtud de (*) y (**) (págs. 183—184), por las fórmulas:

$$\varphi(x \mid y) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\, dx}, \qquad (***)$$

$$\psi(y \mid x) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\, dy}. \qquad (****)$$

Escribimos las fórmulas (*) y (**) en la forma:

$$f(x, y) = f_2(y) \cdot \varphi(x \mid y),$$
$$f(x, y) = f_1(x) \cdot \psi(y \mid x).$$

De donde se deduce que: multiplicando la ley de distribución de una de las componentes por la ley condicional de distribución de la otra componente, hallamos la ley de distribución del sistema de magnitudes aleatorias.

Como toda función diferencial, las funciones diferenciales condicionales poseen las siguientes propiedades:

$$\varphi(x \mid y) \geqslant 0, \quad \int_{-\infty}^{\infty} \varphi(x \mid y)\, dx = 1;$$

$$\psi(y \mid x) \geqslant 0, \quad \int_{-\infty}^{\infty} \psi(y \mid x)\, dy = 1.$$

Ejemplo. La magnitud aleatoria bidimensional (X, Y) está prefijada por la función diferencial

$$f(x, y) = \begin{cases} \frac{1}{\pi r^2} & \text{para} \quad x^2 + y^2 < r^2, \\ 0 & \text{para} \quad x^2 + y^2 > r^2. \end{cases}$$

Hallar las leyes diferenciales condicionales de distribución de las probabilidades de las componentes.

SOLUCION. Hallamos la función diferencial condicional de la componente X por la fórmula (***):

$$\varphi(x \mid y) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\,dx} = \frac{\frac{1}{\pi r^2}}{\frac{1}{\pi r^2} \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx} = \frac{1}{2\sqrt{r^2-y^2}}$$

$$\text{para } |x| < \sqrt{r^2 - y^2}.$$

Dado que $f(x, y) = 0$ cuando $x^2 + y^2 > r^2$, tendremos que $\varphi(x \mid y) = 0$ para

$$|x| > \sqrt{r^2 - y^2}.$$

Utilizando la fórmula (****), análogamente hallamos la función diferencial condicional de la componente Y:

$$\psi(y \mid x) = \begin{cases} \frac{1}{2\sqrt{r^2 - x^2}} & \text{para} \quad |y| < \sqrt{r^2 - x^2}, \\ 0 & \text{para} \quad |y| > \sqrt{r^2 - x^2}. \end{cases}$$

§ 15. Esperanza matemática condicional

La distribución condicional de las probabilidades es una característica importante de la esperanza matemática condicional.

Se llama *esperanza matemática condicional* de una magnitud aleatoria discreta Y para $X = x$ (x es un valor posible determinado de X) el producto de los valores posibles de Y por sus probabilidades condicionales:

$$M(Y \mid X = x) = \sum_{j=1}^{m} y_j p(y_j \mid x). \qquad (*)$$

Para las magnitudes continuas

$$M(Y \mid X=x)=\int_{-\infty}^{\infty} y \cdot \psi(y \mid x)\, dy,$$

donde $\psi(y \mid x)$ es la función diferencial condicional de la magnitud aleatoria Y cuando $X = x$.

Análogamente se determina la esperanza matemática condicional de la magnitud X.

Ejemplo. Una magnitud aleatoria bidimensional discreta está dada por la tabla 5.

Tabla 5

Y \ X	$x_1=1$	$x_2=3$	$x_3=4$	$x_4=8$
$y_1=3$	0,15	0,06	0,25	0,04
$y_2=6$	0,30	0,10	0,03	0,07

Hallar la esperanza matemática condicional de la componente Y si $X = x_1 = 1$.

SOLUCION. Hallamos $p(x_1)$, para ello sumamos las probabilidades de la primera columna de la tabla 5.

$$p(x_1) = 0{,}15 + 0{,}30 = 0{,}45.$$

Hallamos la distribución condicional de las probabilidades de la magnitud Y para $X = x_1 = 1$ (§ 13):

$$p(y_1 \mid x_1)=\frac{p(x_1,\ y_1)}{p(x_1)}=\frac{0{,}15}{0{,}45}=\frac{1}{3};$$

$$p(y_2 \mid x_1)=\frac{p(x_2,\ y_2)}{p(x_1)}=\frac{0{,}30}{0{,}45}=\frac{2}{3}.$$

Hallamos la esperanza matemática condicional buscada por la fórmula (*):

$$M(Y \mid X=x_1)=\sum_{j=1}^{2} y_j p(y_j \mid x_1)=$$

$$=y_1 \cdot p \times (y_1 \mid x_1)+y_2 \cdot p(y_2 \mid x_1)=3 \cdot \frac{1}{3}+6 \cdot \frac{2}{3}=5.$$

§ 16. Magnitudes aleatorias dependientes e independientes

Dos magnitudes aleatorias las hemos denominado independientes cuando la ley de distribución de una de ellas no depende de los valores posibles que tome la otra magnitud. De esta definición se deduce que las distribuciones condicionales de las magnitudes independientes son iguales a sus distribuciones absolutas.

Deducimos las condiciones necesarias y suficientes de independencia de las magnitudes aleatorias.

Teorema. *Para que las magnitudes aleatorias X e Y sean independientes, es necesario y suficiente que la función integral del sistema (X, Y) sea igual al producto de las funciones integrales de las componentes:*

$$F(x, y) = F_1(x) \cdot F_2(y).$$

DEMOSTRACION a) *Necesidad.* Supongamos que X e Y son independientes. En ese caso los sucesos $X < x$ e $Y < y$ son independientes, por lo tanto, la probabilidad de simultaneidad de estos sucesos es igual al producto de sus probabilidades

$$P(X < x, Y < y) = P(X < x) \cdot P(Y < y),$$

o bien

$$F(x, y) = F_1(x) \cdot F_2(y).$$

b) *Suficiencia.* Supongamos que $F(x, y) = F_1(x) \cdot F_2(y)$. De donde

$$P(X < x, Y < y) = P(X < x) \cdot P(Y < y),$$

es decir, la probabilidad de simultaneidad de los sucesos $X < x$ e $Y < y$ es igual al producto de las probabilidades de estos sucesos. Por consiguiente, las magnitudes aleatorias X e Y son independientes.

Corolario. *Para que las magnitudes aleatorias continuas X e Y sean independientes, es necesario y suficiente que la función diferencial del sistema (X, Y) sea igual al producto de las funciones diferenciales de las componentes:*

$$f(x, y) = f_1(x) \cdot f_2(y).$$

DEMOSTRACION. a) *Necesidad.* Supongamos que X e Y sean magnitudes aleatorias continuas independientes. En tal caso (basándose en el teorema anterior)

$$F(x, y) = F_1(x) \cdot F_2(y).$$

Diferenciando esta igualdad por x, y luego por y, tendremos

$$\frac{\partial^2 F}{\partial x\, \partial y} = \frac{\partial F_1}{\partial x} \cdot \frac{\partial F_2}{\partial y}.$$

o bien (por definición de la función diferencial de las magnitudes bidimensional y unidimensional)

$$f(x, y) = f_1(x) \cdot f_2(y).$$

b) *Suficiencia.* Supongamos que

$$f(x, y) = f_1(x) \cdot f_2(y).$$

Integrando esta igualdad por x y por y, obtenemos

$$\int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)\, dx\, dy = \int_{-\infty}^{x} f_1(x)\, dx \int_{-\infty}^{y} f_2(y)\, dy.$$

o bien (§ 8 del cap. XIV y § 3 del cap. XI)

$$F(x, y) = F_1(x) \cdot F_2(y).$$

De donde (basándose en el teorema anterior) deducimos que X e Y son independientes.

Nota. Puesto que las condiciones expuestas antes son necesarias y suficientes, se pueden dar nuevas definiciones de las magnitudes aleatorias independientes:

1) dos magnitudes aleatorias se llaman independientes, si la función integral del sistema de estas magnitudes es igual al producto de las funciones integrales de las componentes;

2) dos magnitudes aleatorias continuas se llaman independientes si la función diferencial del sistema de estas magnitudes es igual al producto de las funciones diferenciales de las componentes.

§ 17. Características numéricas de un sistema de dos magnitudes aleatorias. Momento de correlación. Coeficiente de correlación

Para describir un sistema de dos magnitudes aleatorias, además de las esperanzas matemáticas y las dispersiones de las componentes, también se utilizan otras características, entre las cuales se encuentran el momento de correlación y el coeficiente de correlación.

Se llama *momento de correlación* μ_{xy} de las magnitudes aleatorias X e Y la esperanza matemática del producto de las desviaciones de estas magnitudes:

$$\mu_{xy} = M[(X - M(X))(Y - M(Y))].$$

Para calcular el momento de correlación de las magnitudes discretas se utiliza la fórmula

$$\mu_{xy}=\sum_{i=1}^{n}\sum_{j=1}^{m}[x_i-M(X)][y_j-M(Y)]\,p(x_i, y_j),$$

y para las magnitudes continuas

$$\mu_{xy}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-M(X)][y-M(Y)]\,f(x, y)\,dx\,dy.$$

El momento de correlación sirve para caracterizar el enlace entre las magnitudes X e Y. Como se demostrará más adelante, el momento de correlación es igual a cero, si X e Y son independientes; por lo tanto, si el momento de correlación es distinto de cero, X e Y son magnitudes aleatorias dependientes.

Teorema. *El momento de correlación de dos magnitudes aleatorias independientes X e Y es igual a cero.*

DEMOSTRACIÓN. Ya que X e Y son magnitudes aleatorias independientes, sus desviaciones $X-M(X)$ e $Y-M(Y)$ también son independientes. Utilizando las propiedades de la esperanza matemática (la esperanza matemática del producto de magnitudes aleatorias independientes es igual al producto de las esperanzas matemáticas de los factores) y de la desviación (la esperanza matemática de la desivación es igual a cero), obtenemos

$$\mu_{xy}=M[(X-M(X))\cdot(Y-M(Y))]=$$
$$=M[X-M(X)]\cdot M[Y-M(Y)]=0.$$

De la definición de momento de correlación se deduce que éste tiene una dimensión igual al producto de las dimensiones de las magnitudes X e Y. En otras palabras, la magnitud del momento de correlación depende de las unidades de medición de las magnitudes aleatorias. Por este motivo, para dos magnitudes idénticas la dimensión del momento de correlación tendrá distintos valores según las unidades utilizadas al medir las magnitudes.

Supongamos, por ejemplo, que X e Y han sido medidos en centímetros y $\mu_{xy}=2$ cm^2, si X e Y se miden en milímetros, $\mu_{xy}=200$ mm^2. Esta particularidad del momento de correlación es una deficiencia de esa característica numé-

rica, puesto que la comparación de los momentos de correlación de distintos sistemas de magnitudes aleatorias se hace dificultoso. Para evitar este inconveniente se introduce una nueva característica numérica, o sea, el coeficiente de correlación.

Se llama *coeficiente de correlación* r_{xy} de las magnitudes aleatorias X e Y la relación entre el momento de correlación y el producto de las desviaciones medias cuadráticas de estas magnitudes:

$$r_{xy} = \frac{xy}{\sigma_x \sigma_y} .$$

Ya que la dimensión de μ_{xy} es igual al producto de las dimensiones de las magnitudes X e Y, σ_x tiene la dimensión de la magnitud X, σ_y, la dimensión de la magnitud Y (cap. VIII, § 7), tendremos que r_{xy} es una magnitud adimensional. De este modo, la magnitud del coeficiente de correlación no depende de la selección de las unidades de medición de las magnitudes aleatorias. En esto reside la ventaja del coeficiente de correlación respecto del momento de correlación.

Evidentemente, el coeficiente de correlación de magnitudes aleatorias independientes es igual a cero (puesto que $\mu_{xy} = 0$).

Nota. En muchos problemas de la teoría de las probabilidades conviene considerar, en lugar de la magnitud aleatoria X, la magnitud normada X' que se determina como relación de la desviación a la desviación cuadrática media:

$$X' = \frac{X - M(X)}{\sigma_x} .$$

La magnitud normada tiene una esperanza matemática igual a cero y la dispersión, igual a la unidad. En efecto, utilizando las propiedades de la esperanza matemática y de la dispersión, tendremos:

$$M(X') = M\left(\frac{X - M(X)}{\sigma_x}\right) = \frac{1}{\sigma_x} \cdot M[X - M(X)] = \frac{1}{\sigma_x} \cdot 0 = 0;$$

$$D(X') = D\left(\frac{X - M(X)}{\sigma_x}\right) = \frac{1}{\sigma_x^2} \cdot D(X - M(X)) = \frac{D(X)}{\sigma_x^2} = 1.$$

Se comprueba fácilmente que el coeficiente de correlación r_{xy} es igual al momento de correlación de las magnitudes normadas X' e Y':

$$r_{xy} = \frac{M[(X - M(X)) \cdot (Y - M(Y))]}{\sigma_x \sigma_y} = M\left[\frac{X - M(X)}{\sigma_x} \frac{Y - M(Y)}{\sigma_y}\right] = M(X' \cdot Y') = \mu_{x'y'} .$$

§ 18. Correlación y dependencia de magnitudes aleatorias

Dos magnitudes aleatorias X e Y se llaman *correlacionadas*, si su momento de correlación (o, lo que es igual, el coeficiente de correlación) es distinto de cero; X e Y se llaman magnitudes *no correlacionadas*, si su momento de correlación es igual a cero.

Dos magnitudes correlacionadas también son dependientes. En efecto, admitiendo lo contrario, debemos deducir que $\mu_{xy} = 0$, lo que no puede ser ya que por la condición para las magnitudes correlacionadas $\mu_{xy} \neq 0$.

La enunciación inversa no siempre es justa, o sea, si dos magnitudes son dependientes, ellas pueden ser tanto correlacionadas, como no correlacionadas. En otras palabras, el momento de correlación de dos magnitudes dependientes puede ser distinto de cero, pero también puede igualarse a cero.

Comprobemos con un ejemplo que dos magnitudes dependientes pueden ser no correlacionadas.

Ejemplo. La magnitud aleatoria bidimensional (X, Y) está prefijada por la función diferencial

$$f(x, y) = \frac{1}{6\pi} \text{ dentro de la elipse } \frac{x^2}{9} + \frac{y^2}{4} = 1,$$

$$f(x, y) = 0 \text{ fuera de esta elipse.}$$

Demostrar que X e Y son magnitudes no correlacionadas dependientes.

SOLUCIÓN. Utilizamos las funciones diferenciales de X e Y antes calculadas (§ 12):

$f_1(x) = \frac{2}{9\pi}\sqrt{9-x^2}$, $f_2(y) = \frac{1}{2\pi}\sqrt{4-y^2}$ dentro de la elipse dada y $f_1(x) = 0$, $f_2(y) = 0$ fuera de ella.

Puesto que $f(x, y) \neq f_1(x) \cdot f_2(y)$, tendremos que X e Y son magnitudes dependientes (§ 16).

Para demostrar la no correlación de X e Y es suficiente cerciorarse de que $\mu_{xy} = 0$.

Hallamos el momento de correlación por la fórmula (§ 17):

$$\mu_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [x - M(X)]\,[y - M(Y)]\, f(x, y)\, dx\, dy$$

Dado que la función diferencial $f_1(x)$ es simétrica respecto al eje OY, entonces $M(x) = 0$; análogamente $M(Y) = 0$, en virtud de la simetría de $f_2(y)$ respecto al eje OX. Por lo tanto,

$$\mu_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x, y)\, dx\, dy.$$

Sacando el factor constante $f(x, y)$ fuera de la integral, obtenemos

$$\mu_{xy} = f(x, y) \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} x\, dx \right) dy.$$

La integral entre paréntesis es igual a cero (la función subintegral es impar, los límites de integración son simétricos con respecto al origen de coordenadas), por lo tanto, $\mu_{xy} = 0$, es decir, las magnitudes aleatorias dependientes X e Y son no correlacionadas.

De este modo, de la correlación de dos magnitudes aleatorias se deduce su dependencia, pero de la dependencia aún no se deduce la correlación. De la independencia de dos magnitudes se deduce su no correlación, pero de la no correlación no se puede deducir aún la independencia de estas magnitudes.

Sin embargo, cabe hacer notar que de la no correlación de magnitudes distribuidas normalmente se desprende la independencia de las mismas. Esta tesis será demostrada en el párrafo siguiente.

§ 19. Ley normal de distribución en el plano

Frecuentemente en la práctica se tropieza con magnitudes aleatorias bidimensionales, distribuidas normalmente.

Se llama *ley normal de distribución en el plano* la distribución de las probabilidades de una magnitud aleatoria bidimensional (X, Y), cuando

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2}} \times$$

$$\times e^{-\frac{1}{2(1-r_{xy}^2)}\left[\frac{(x-a_1)^2}{\sigma_x^2} + \frac{(y-a_2)^2}{\sigma_y^2} - 2r_{xy}\frac{x-a_1}{\sigma_x}\cdot\frac{y-a_2}{\sigma_y}\right]}. \qquad (*)$$

Como vemos, la ley normal en el plano se determina por cinco parámetros: a_1, a_2, τ_x, σ_y y r_{xy}. Se puede demostrar que estos parámetros tienen los siguientes sentidos probabilísticos:

a_1, a_2 son las esperanzas matemáticas,

σ_x, σ_y son desviaciones cuadráticas medias,

r_{xy} es el coeficiente de correlación de las magnitudes X e Y.

Nos cercioramos de que si las componentes de una magnitud aleatoria bidimensional normalmente distribuida son no correlacionadas, entonces también son independientes. En efecto, supongamos que X e Y son no correlacionadas. En ese caso, si en la fórmula (*) admitimos que $r_{xy} = 0$, obtenemos

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \cdot e^{-\frac{1}{2}\left[\frac{(x-a_1)^2}{\sigma_x^2} + \frac{(y-a_2)^2}{\sigma_y^2}\right]} =$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} \cdot e^{-\frac{(x-a_1)^2}{2\sigma_x^2}} \cdot \frac{1}{\sigma_y\sqrt{2\pi}} \cdot e^{-\frac{(y-a_2)^2}{2\sigma_y^2}} = f_1(x) \cdot f_2(y).$$

De este modo, si las componentes de una magnitud aleatoria normalmente distribuida son no correlacionadas, la función diferencial del sistema es igual al producto de las funciones diferenciales de las componentes, de donde, precisamente, se deduce la independencia de las componentes (§ 16). Es cierta también la tesis inversa (§ 18).

Así pues, para las componentes normalmente distribuidas de una magnitud aleatoria bidimensional los conceptos de independencia y de no correlación son equivalentes.

Problemas

1. Hallar las leyes de distribución de las componentes de una magnitud aleatoria discreta prefijada por la ley de distribución

Y \ X	x_1	x_2	x_3
y_1	0,12	0,18	0,10
y_2	0,10	0,11	0,39

Respuesta

X	x_1	x_2	x_3	Y	y_1	y_2
p	0,22	0,29	0,49	p	0,40	0,60

2. Hallar la probabilidad de que la componente Y de una magnitud aleatoria bidimensional tome el valor $X < \frac{1}{2}$ y en ese caso, la componente Y tome el valor $Y < \frac{1}{3}$, si se conoce la función integral del sistema

$$F(x, y) = \left(\frac{1}{\pi}\operatorname{arctg} 2x + \frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg} 3y + \frac{1}{2}\right).$$

Respuesta $P\left(X < \frac{1}{2},\ Y < \frac{1}{3}\right) = \frac{9}{16}$.

3. Hallar la probabilidad de que el punto aleatorio (X, Y) caiga en el rectángulo, limitado por las rectas $x = \frac{\pi}{4}$, $x = \frac{\pi}{2}$, $y = \frac{\pi}{6}$, $y = \frac{\pi}{3}$, si se conoce la función integral

$$F(x, y) = \operatorname{sen} x \operatorname{sen} y \left(0 \leqslant x \leqslant \frac{\pi}{2},\ 0 \leqslant y \leqslant \frac{\pi}{2}\right).$$

Respuesta $P\left(\frac{\pi}{4} < X < \frac{\pi}{2},\ \frac{\pi}{6} < Y < \frac{\pi}{3}\right) = 0{,}11$.

4. Hallar la función diferencial de un sistema de dos magnitudes aleatorias mediante la función integral

$$F(x, y) = (1 - e^{-2x})(1 - e^{-3y})\ (x \geqslant 0,\ y \geqslant 0).$$

Respuesta $f(x, y) = \frac{\partial^2 F}{\partial x\, \partial y} = 6e^{-(2x+3y)}$.

5. En el interior del rectángulo limitado por las rectas $x = 0$, $x = \frac{\pi}{2}$, $y = 0$, $y = \frac{\pi}{2}$, la función diferencial de un sistema de dos magnitudes aleatorias $f(x, y) = C \operatorname{sen}(x + y)$, fuera del rectángulo, $f(x, y) = 0$. Hallar: a) la magnitud C, b) la función integral del sistema.

Respuesta a) $C = 0{,}5$; b) $F(x, y) = 0{,}5[\operatorname{sen} x + \operatorname{sen} y - \operatorname{sen}(x + y)] \times$

$$\times \left(0 \leqslant x \leqslant \frac{\pi}{2},\ 0 < y < \frac{\pi}{2}\right).$$

6. Un sistema de dos magnitudes aleatorias está distribuido uniformemente: en un rectángulo limitado por las rectas $x = 4$, $x = 6$, $y = 10$, $y = 15$, la función diferencial conserva un valor constante,

y fuera de este rectángulo ella es igual a cero. Hallar: a) la función diferencial, b) la función integral del sistema.

Respuesta a) $f(x,\ y) = \begin{cases} 0{,}1 \text{ en el interior del rectángulo,} \\ 0 \text{ fuera del rectángulo;} \end{cases}$

$$\text{b) } F(x,\ y) = \frac{(x-4)(y-10)}{10}.$$

7. La función diferencial de un sistema de dos magnitudes aleatorias $f(x,\ y) = \dfrac{C}{(4+x^2)(9+y^2)}$. Hallar: a) la magnitud C, b) la función integral del sistema.

Respuesta a) $C = \dfrac{6}{\pi^2}$;

$$\text{b) } F(x,\ y) = \left(\frac{1}{\pi}\operatorname{arctg}\frac{x}{2} + \frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg}\frac{y}{3} + \frac{1}{2}\right).$$

8. Una magnitud aleatoria bidimensional se prefija por la función diferencial

$$f(x,\ y) = \frac{3\sqrt{3}}{\pi}\, e^{-4x^2-6xy-9y^2}.$$

Hallar las leyes condicionales de distribución de las componentes

Respuesta $\varphi(x \mid y) = \dfrac{2}{\sqrt{\pi}}\, e^{-\left(2x+\frac{3}{2}y\right)^2}$.

$$\psi(x \mid y) = \frac{3}{\sqrt{\pi}}\, e^{-(x+3y)^2}.$$

Parte tercera

Elementos de estadística matemática

Capítulo quince

METODO MUESTRAL

§ 1. Objetivo de la estadística matemática

La determinación de las leyes, a las que obedecen los fenómenos aleatorios de masas, se basa en el estudio de los datos estadísticos, o sea, resultados de observaciones. El primer objetivo de la estadística matemática es indicar los métodos de recogida y agrupamiento (si los datos son muchísimos) de los datos estadísticos.

El segundo objetivo de la estadística matemática es la elaboración de los métodos de análisis de datos estadísticos, en función de los propósitos de la investigación.

El estudio de unos u otros fenómenos por los métodos de la estadística matemática sirve como medio de resolución de muchos problemas, presentados por la ciencia y la práctica (la organización correcta del proceso tecnológico, la planificación más conveniente, etc.).

Así, el objetivo de la estadística matemática es la creación de los métodos de recogida y elaboración de datos estadísticos para obtener conclusiones científicas y prácticas.

§ 2. Breve información histórica

La estadística matemática surgió (siglo XVII) y se formó paralelamente con la teoría de las probabilidades. El desarrollo ulterior de la estadística matemática (segunda mitad del siglo XIX y comienzo del siglo XX) se debe, en primer término, a P. L. Chebishev, A. A. Markov, A. M. Liapunov, así como K. Gauss, A. Quételet, F. Galton, K. Pearson, etc.

Las mayores aportaciones efectuadas a la estadística matemática en el siglo XX han sido las de los matemáticos

soviéticos (V. I. Romanovsky, E. E. Sluszky, A. N. Kolmogorov, N. V. Smirnov), así como los ingleses (Student, R. Fisher, E. Pearson) y los norteamericanos (J. Neyman, A. Wald).

§ 3. Conjunto general y muestral

Supongamos que se quiere estudiar un conjunto (población) de objetos homogéneos respecto a cierto *índice cualitativo* o *cuantitativo* que caracteriza estos objetos. Por ejemplo, si se tiene un lote de piezas, como índice cualitativo puede servir el standard de la pieza y como cuantitativo, la dimensión controlable de la pieza.

A veces se realiza una investigación total, es decir, se examina *cada* uno de los objetos del conjunto respecto al índice que interesa. En la práctica, sin embargo, la investigación total se practica con relativa rareza. Por ejemplo, si el conjunto contiene un número muy grande de objetos, físicamente es imposible realizar un examen total. Si el examen del objeto está vinculado con su destrucción o requiere grandes gastos materiales, prácticamente no tiene sentido efectuar tal investigación total. En estos casos se escogen fortuitamente del total un número limitado de objetos y se someten éstos al estudio.

Se llama *conjunto muestral*, o simplemente muestra, conjunto de objetos tomados fortuitamente.

Se llama *conjunto general* el conjunto de objetos, de los cuales se hace muestreo.

Se llama *volumen* del conjunto (muestral o general) el número de objetos de ese conjunto. Por ejemplo, si de 1000 piezas se escogen para el examen 100 piezas, el volumen del conjunto general es $N = 1000$ y el volumen de la muestra $n = 100$.

Nota. De ordinario, el conjunto general contiene un número finito de objetos. Sin embargo, si ese número es bastante grande, a veces para simplificar los cálculos o para facilitar las deducciones teóricas, se supone que el conjunto general se compone de una infinidad de objetos. Esta suposición se justifica, ya que el aumento del volumen del conjunto general (un volumen suficientemente grande), prácticamente no se manifiesta en los resultados de la elaboración de los datos de la muestra.

§ 4. Muestras repetida y única. Muestra representativa

Al componer la muestra se puede proceder de dos formas: después que el objeto ha sido escogido y se le ha observado, puede ser reintegrado o no al conjunto general (a la totalidad). De acuerdo con lo dicho las muestras se dividen en repetidas y únicas.

La muestra se llama *repetida*, cuando el objeto escogido (antes de tomar el siguiente) se reintegra al conjunto general.

La muestra se llama *única*, cuando el objeto escogido no se restituye al conjunto general.

En la práctica se utiliza generalmente la selección aleatoria única.

Para que por los datos de la muestra se pueda juzgar con bastante certeza el índice que nos interesa del conjunto general, es necesario que el objeto de la muestra lo represente correctamente. Este requisito se formula brevemente así: la muestra debe ser *representativa*.

En virtud de la ley de los grandes números se puede afirmar que, la muestra será representativa, si se realiza fortuitamente: cada objeto de la muestra se escoge al azar del conjunto general, cuando todos los objetos tienen igual probabilidad de caer en la muestra.

Si el volumen del conjunto general es suficientemente grande, y la muestra constituye solamente una parte ínfima de este conjunto, la diferencia entre las muestras repetida y única se elimina; en el caso límite, cuando se analiza un conjunto general infinito, en tanto que la muestra tiene un volumen finito, esa diferencia desaparece.

§ 5. Métodos de selección

En la práctica se utilizan distintos métodos de selección. Estos métodos se pueden dividir, principalmente, en dos tipos:

1. La selección que no requiere el desmembramiento del conjunto general, o totalidad, en partes; aquí se distinguen:

a) la selección única aleatoria simple;

b) la selección repetida aleatoria simple.

2. La selección, para la cual el conjunto general (el total) se divide en partes; aquí se distinguen:

a) la selección típica;

b) la selección mecánica;
c) la selección en serie.

Una selección se llama *aleatoria simple*, cuando los objetos se extraen uno a la vez del total. La selección simple puede realizarse de distintos modos. Por ejemplo, para extraer n objetos de un conjunto general de volumen N se procede así: se escriben sobre tarjetas números desde 1 hasta N y se mezclan bien, se extrae una tarjeta al azar; el objeto que tiene igual número que la tarjeta extraída se somete a examen; a continuación la tarjeta se reintegra al paquete y se repite el proceso, es decir, se mezclan las tarjetas y se extrae al azar una de ellas, etc. Así se procede n veces; en suma se obtiene una muestra repetida aleatoria simple de volumen n.

Si las cartas extraídas no se reintegran al paquete, la muestra será única aleatoria simple.

Cuando el volumen del conjunto general es grande, el proceso descrito resulta muy dificultoso. En ese caso, se utilizan tablas de «números aleatorios», en las cuales los números están dispuestos en orden aleatorio. Para seleccionar, por ejemplo, 50 objetos, de un conjunto general numerado, se abre cualquier página de la tabla de números aleatorios y se escribe en orden 50 números; en la muestra caen aquellos objetos cuyos números coinciden con los aleatorios escritos. Si resultase que un número aleatorio de la tabla es mayor que el número N, éste se deja pasar. Al realizar una muestra única de números aleatorios de la tabla, el número antes encontrado también se omite.

La selección se llama *típica* cuando los objetos no se toman del conjunto general, sino de cada una de su parte «típica». Por ejemplo, si las piezas se producen en varias máquinas, la selección no se realiza de todo el conjunto de piezas, elaboradas en todas las máquinas, sino de la producción de cada máquina por separado. La selección típica se utiliza cuando el índice que se examina oscila en las distintas partes típicas del conjunto general. Por ejemplo, si la producción se prepara en varias máquinas, entre las cuales hay más o menos desgastadas, aquí es conveniente la selección típica.

La selección se llama *mecánica*, cuando el conjunto general se divide «mecánicamente» en tantos grupos, como objetos deben entrar en la muestra y de cada grupo se toma un objeto.

Por ejemplo, si hay que seleccionar el 20% de piezas producidas por una máquina, se toma cada quinta pieza; si hay que seleccionar el 5% de piezas, se toma una de cada veinte piezas, etc.

Cabe hacer notar que a veces la selección mecánica puede no garantizar una muestra representativa. Por ejemplo, si se toma cada vigésimo eje torneado; además, inmediatamente después de la selección se cambia la cuchilla, resultan seleccionados todos los ejes torneados con cuchillas desafiladas. En ese caso hay que evitar la coincidencia del ritmo de selección con el ritmo de sustitución de la cuchilla, para ello, hay que tomar, digamos, cada décimo eje de las veinte torneadas.

La selección se llama *en serie*, cuando los objetos se seleccionan del conjunto general por «series», y no de uno en uno; estas series se someten a un examen completo. Por ejemplo, si los artículos se producen por un grupo grande de máquinas automáticas, se someten a un examen completo solamente la producción de algunas máquinas. La selección en serie se utiliza cuando el índice a investigar oscila poco en distintas series.

Conviene subrayar que en la práctica, frecuentemente, se utiliza la selección combinada, en la que se reúnen los métodos antes indicados.

Por ejemplo, a veces se divide el conjunto general en series de igual volumen, después por selección aleatoria simple se toman varias series y, por último, de cada serie por selección aleatoria simple se extraen objetos separados.

§ 6. Distribución estadística de la muestra

Supongamos que del conjunto general se ha extraído una muestra; además, x_1 se observó n_1 veces, x_2, n_2 veces, x_k, n_k veces y $\sum n_i = n$ es el volumen de la muestra. Los valores observados de x_i se llaman *variantes*, y la sucesión de variantes escritas en orden creciente, serie de variación. Los números de observaciones se llaman *frecuencias*, y su relación al volumen de la muestra $\frac{n_i}{n} = W_i$, *frecuencias relativas*.

Se llama *distribución estadística de la muestra* la enumeración de variantes y sus correspondientes frecuencias o frecuencias relativas. La distribución estadística se puede

prefijar también en forma de sucesión de intervalos y sus correspondientes frecuencias (como frecuencia correspondiente al intervalo, se toma la suma de frecuencias que caen en ese intervalo).

Cabe recordar que en la teoría de las probabilidades por *distribución* se entiende la correspondencia entre los valores posibles de una magnitud aleatoria y sus probabilidades, en tanto que en la estadística matemática, la correspondencia entre las variantes observadas y sus frecuencias, o frecuencias relativas.

Ejemplo. Dada la distribución de frecuencias de la muestra de volumen = 20:

x_i	2	6	12
n_i	3	10	7.

Escribir la distribución de frecuencias relativas.

SOLUCION. Hallemos las frecuencias relativas; para ello dividimos las frecuencias por el volumen de la muestra:

$$W_1=\frac{3}{20}=0{,}15,\quad W_2=\frac{10}{20}=0{,}50,\quad W_3=\frac{7}{20}=0{,}35.$$

Escribimos la distribución de las frecuencias relativas:

x_i	2	6	12
W_i	0,15	0,5	0,35.

Verificación: $0{,}15+0{,}5+0{,}35=1$.

§ 7. Función empírica de distribución

Supongamos que se conoce la distribución estadística de frecuencias del carácter cuantitativo de X. Introducimos las designaciones:

n_x, el número de observaciones, durante las cuales se observó un valor del carácter, menor que x,

n, el número total de observaciones (volumen de la muestra).

Es evidente que la frecuencia relativa del suceso $X<x$ es igual a $\frac{n_x}{n}$. Si x varía, en general, variará también la frecuencia relativa, es decir, la frecuencia relativa $\frac{n_x}{n}$ es una función de x. Dado que esta función se halla empíricamente (experimentalmente), se llama empírica.

Se llama *función empírica de distribución* (función de distribución de la muestra) a la función $F^*(x)$ que determina para cada valor de x la frecuencia relativa del suceso $X < x$.

De este modo, por definición

$$F^*(x) = \frac{n_x}{n},$$

donde n_x es el número de las variantes menores que x,
n es el volumen de la muestra.

Por consiguiente, para hallar, por ejemplo, $F^*(x_2)$, el número de las variantes menores que x_2, hay de dividirlo por el volumen de la muestra:

$$F^*(x_2) = \frac{n_{x_2}}{n}.$$

A diferencia de la función empírica de distribución de la muestra, la función integral $F(x)$ de distribución del conjunto general se llama *función teórica de distribución*. La diferencia entre las funciones empírica y teórica está en que la función teórica $F(x)$ determina la probabilidad del suceso $X < x$, en tanto que la función empírica $F^*(x)$ determina la frecuencia relativa de ese suceso. Del teorema de Bernoulli se deduce que la frecuencia relativa del suceso $X < x$, es decir, $F^*(x)$ tiende en probabilidad a la probabilidad $F(x)$ de este suceso. En otras palabras, los números $F^*(x)$ y $F(x)$ se diferencian poco entre sí. Ya de aquí se deduce la conveniencia de utilizar la función empírica de distribución de la muestra, para una representación aproximada de la función teórica (integral) de distribución del conjunto general.

Esta deducción se confirma con que $F^*(x)$ posee todas las propiedades de $F(x)$. En efecto, de la definición de la función $F^*(x)$ se desprenden sus siguientes propiedades:

1) los valores de la función empírica corresponden al segmento $[0, 1]$;

2) $F^*(x)$ es una función no decreciente;

3) si x_1 es la variante menor, tendremos que $F^*(x) = 0$ cuando $x \leqslant x_1$;
si x_k es la variante mayor, $F^*(x) = 1$ cuando $x > x_k$.

De este modo, la función empírica de distribución de una muestra sirve para estimar la función teórica de distribución del conjunto general.

Ejemplo. Formar la función empírica por la distribución dada de la muestra

variantes x_i	2	6	10
frecuencias n_i	12	18	30.

SOLUCION. Hallamos el volumen de la muestra: $12 + 18 + 30 = 60$.

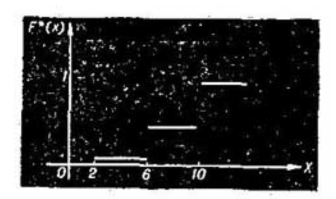


Fig. 19.

La variante mínima es igual a 2, por lo tanto,

$$F^*(x) = 0, \quad \text{cuando} \quad x \leqslant 2.$$

El valor de $X < 6$ y, precisamente, $x_1 = 2$ se observó 12 veces; por lo tanto

$$F^*(x) = \frac{12}{60} = 0{,}2 \quad \text{para} \quad 2 < x \leqslant 6.$$

El valor de $X < 10$ y, precisamente, $x_1 = 2$ y $x_2 = 6$ se observaron $12 + 18 = 30$ veces; en consecuencia

$$F^*(x) = \frac{30}{60} = 0{,}5 \quad \text{para} \quad 6 < x \leqslant 10.$$

Ya que $x = 10$ es la variante máxima, tenemos que

$$F^*(x) = 1 \quad \text{cuando} \quad x > 10.$$

La función empírica buscada es

$$F^*(x) = \begin{cases} 0 & \text{para} \quad x \leqslant 2, \\ 0{,}2 & \text{para} \quad 2 < x \leqslant 6, \\ 0{,}5 & \text{para} \quad 6 < x \leqslant 10, \\ 1 & \text{para} \quad x > 10. \end{cases}$$

La gráfica de esta función está representada en la fig. 19.

§ 8. Polígono e histograma

Para claridad se construyen distintas gráficas de la distribución estadística y, en particular, el polígono y el histograma.

Se llama *polígono de frecuencias* la línea quebrada, cuyos segmentos están unidos por los puntos (x_1, n_1), (x_2, n_2), ..., (x_k, n_k). Para construir el polígono de frecuencias sobre el eje de abscisas se llevan las variantes x_i y sobre el

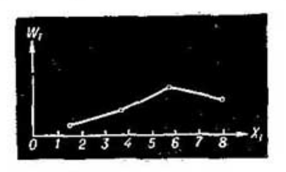


Fig. 20.

eje de ordenadas, sus respectivas frecuencias n_i. Los puntos (x_i, n_i) se unen por segmentos de rectas y se obtiene el polígono de frecuencias.

Se llama *polígono de frecuencias relativas* la línea quebrada, cuyos segmentos se unen por los puntos (x_1, W_1), (x_2, W_2), . . ., (x_k, W_k). Para construir el polígono de frecuencias relativas sobre el eje de abscisas se llevan las variantes x_i, y sobre el eje de ordenadas, sus respectivas frecuencias relativas W_i. Los puntos (x_i, W_i) se unen por segmentos de rectas y se obtiene el polígono de frecuencias relativas.

En la fig. 20 se muestra el polígono de frecuencias relativas de la distribución siguiente:

X	1,5	3.5	5,5	7,5
W	0,1	0,2	0,4	0,3.

En el caso de un criterio o carácter continuo conviene construir el histograma, para lo cual el intervalo, en el que están comprendidos todos los valores observables del criterio, se divide en varios intervalos parciales de longitud h y para cada intervalo parcial se halla n_i, o sea, la suma de las frecuencias de las variantes que caen en el i-ésimo intervalo.

Se llama *histograma de frecuencias* la figura compuesta de rectángulos, cuyas bases son los intervalos parciales de longitud h y las alturas, iguales a la relación $\frac{n_i}{h}$ (densidad de frecuencia).

Para construir el histograma de frecuencias sobre el eje de abscisas se llevan los intervalos parciales y sobre ellos

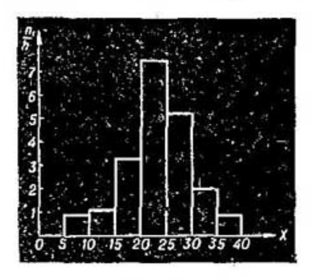


Fig. 21.

se trazan segmentos paralelos al eje de abscisas a la distancia $\frac{n_i}{h}$.

El área del i-ésimo rectángulo es igual a $h \cdot \frac{n_i}{h} = n_i$, o sea, a la suma de frecuencias de las variantes del i-ésimo intervalo; por lo tanto, *el área del histograma de frecuencias es igual a la suma de todas las frecuencias, es decir, al volumen de la muestra*.

En la fig. 21 se muestra el histograma de frecuencias de distribución del volumen $n = 100$, dado en la tabla 6.

Se llama *histograma de frecuencias relativas* la figura en escalera (escalonada) compuesta de rectángulos, cuyas bases son los intervalos parciales de longitud h, mientras que sus alturas son iguales a la relación $\frac{W_i}{h}$ (densidad de frecuencia relativa).

Para construir el histograma de frecuencias relativas sobre el eje de abscisas se llevan los intervalos parciales y sobre ellos se trazan segmentos paralelos al eje de abscisas

Intervalo parcial de longitud $h = 5$	Suma de frecuencias de las variantes de intervalo parcial n_i	Densidad de frencuencia $\frac{n_i}{h}$
5—10	4	0,8
10—15	6	1,2
15—20	16	3,2
20—25	36	7,2
25—30	24	4,8
30—35	10	2,0
35—40	4	0,8

a la distancia $\frac{W_i}{h}$. El área del i-ésimo rectángulo es igual a h. $\frac{W_i}{h} = W_i$, o sea, a la frecuencia relativa de las variantes pertenecientes al i-ésimo intervalo. Por lo tanto, *el área del histograma de frecuencias relativas es igual a la suma de todas las frecuencias relativas, es decir, a la unidad.*

Problemas

1. Construir la gráfica de la función empírica de distribución

x_i	5	7	10	15
n_i	2	3	8	7.

2. Construir los polígonos de frecuencias y de frecuencias relativas de distribución

x_i	1	3	5	7	9
n_i	10	15	30	33	12.

3. Construir los histogramas de frecuencias y de frecuencias relativas de distribución (en la primera columna se indica el intervalo parcial, en la segunda, la suma de frecuencias de las variantes de intervalo parcial):

2—5	9
5—8	10
8—11	25
11—14	6.

Capítulo diez y seis

ESTIMACIONES ESTADISTICAS DE LOS PARAMETROS DE UNA DISTRIBUCION

§ 1. Estimaciones estadísticas de los parámetros de una distribución

Supongamos que se quiere estudiar el carácter cuantitativo de un conjunto general. Admitamos que de las consideraciones teóricas se haya logrado establecer, precisamente qué distribución tiene el carácter. Naturalmente surge el problema de estimar los parámetros que determinan esta distribución. Por ejemplo, si se conoce previamente que el carácter estudiado está distribuido normalmente en el conjunto general, hay que estimar (hallar aproximadamente) la esperanza matemática y la desviación cuadrática media, ya que estos dos parámetros determinan completamente la distribución normal; si se puede considerar que el carácter tiene, por ejemplo, la distribución de Poisson, es necesario estimar el parámetro λ que determina esta distribución.

Generalmente el investigador dispone solamente de los datos de la muestra, por ejemplo, los valores del carácter cuantitativo $x_1, x_2, \ldots, x_n$, obtenidos como resultado de n observaciones (aquí y en adelante las observaciones se suponen independientes). Mediante estos datos se expresa el parámetro a estimar.

Considerando $x_1, x_2, \ldots, x_n$ como magnitudes aleatorias independientes de $X_1, X_2, \ldots, X_n$, podemos decir que hallar la estimación estadística de un parámetro desconocido de una distribución teórica significa hallar la función de las magnitudes aleatorias a observar, la que da un valor aproximado del parámetro estimado. Por ejemplo, como se demostrará más adelante, para estimar la esperanza matemática de distribución normal se utiliza la función (media aritmética de los valores observados del carácter):

$$\overline{X}=\frac{X_1+X_2+\ldots+X_n}{n}.$$

Así pues, se llama estimación estadística de un parámetro desconocido de una distribución normal la función de las magnitudes aleatorias observadas.

§ 2. Estimaciones no desviadas, eficaces y valederas

Para que las estimaciones estadísticas den «buenas» aproximaciones de los parámetros estimados, ellas deben satisfacer determinados requisitos. A continuación se indican estas exigencias.

Dado que Θ^* es la estimación estadística de un parámetro desconocido Θ de una distribución teórica. Admitamos que mediante la muestra de volumen n está hallada estimación Θ_1^*. Repetimos el experimento, es decir, extraemos del conjunto general otra muestra de igual volumen y por sus datos obtenemos la estimación Θ_2^*. Reiterando la prueba varias veces, obtenemos los números Θ_1^*, Θ_2^*, . . ., Θ_k^* que, en general, serán diferentes entre sí. Por consiguiente, la estimación Θ^* se puede considerar como una magnitud aleatoria, mientras que los números Θ_1^*, Θ_2^*, . . ., Θ_k^*, como sus valores posibles.

Supongamos que la estimación Θ^* da un valor aproximado de Θ con exceso; en tal caso, cada número Θ_i^* ($i = 1, 2, \ldots, k$), hallado según los datos de las muestras, será mayor que el valor real de Θ. Evidentemente, en este caso la esperanza matemática (valor medio) de la magnitud aleatoria Θ^* también será mayor que Θ, es decir, $M(\Theta^*) > \Theta$. Está claro que si Θ^* da una estimación con defecto, tendremos que $M(\Theta^*) < \Theta$.

De este modo, el empleo de la estimación estadística, cuya esperanza matemática no es igual al parámetro a estimar, daría lugar a errores sistemáticos (del mismo signo). Por este motivo es natural exigir que la esperanza matemática de la estimación Θ^* sea igual al parámetro que se estima. A pesar de que este requisito no elimina los errores (unos valores de Θ^* son mayores y otros son menores que Θ), sin embargo con igual frecuencia se tropezarán con errores de distintos signos. En otras palabras, el cumplimiento de $M(\Theta^*) = \Theta$ garantiza contra la obtención de errores sistemáticos.

La estimación estadística Θ^* cuya esperanza matemática es igual al parámetro que se estima Θ para todo volumen de la muestra, es decir,

$$M(\Theta^*) = \Theta,$$

se llama *no desviada*.

La estimación cuya esperanza matemática no es igual al parámetro que se estima, se llama *desviada*.

Empero sería erróneo considerar que la estimación no desviada siempre da una buena aproximación del parámetro que se estima. En efecto, los valores posibles de Θ^* pueden ser fuertemente dispersos en torno a su valor medio, es decir, la dispersión $D(\Theta^*)$ puede ser considerable. En este caso, la estimación hallada por los datos de una muestra, por ejemplo, Θ_1^*, puede resultar muy alejada del valor medio $\overline{\Theta^*}$, y por lo tanto, también del propio parámetro estimado Θ; tomando Θ_1^* como valor aproximado de Θ, cometeríamos un gran error. Si se necesita que la dispersión Θ^* sea pequeña, se excluye la posibilidad de cometer un gran error. Por esta causa la estimación estadística debe satisfacer el requisito de *eficacia*.

La estimación estadística se llama *eficaz* cuando tiene la dispersión mínima posible (para un volumen dado de la muestra n).

Al considerar muestras de gran volumen (¡n es grande!) la estimación estadística debe satisfacer el requisito de *validez*.

La estimación estadística se llama *valedera* cuando tiende respecto a la probabilidad al parámetro que se estima para $n \to \infty$. Por ejemplo, si la dispersión de la estimación no desviada tiende a cero para $n \to \infty$, esta estimación resulta precisamente valedera.

§ 3. Media general

Supongamos que se estudia el carácter cuantitativo X de un conjunto general discreto.

Se llama *media general* $\overline{x}_g$ la media aritmética de los valores del carácter del conjunto general.

Si todos los valores $x_1, x_2, \ldots, x_N$ del carácter del conjunto general de volumen N son *distintos*, tendremos que

$$\overline{x}_g = \frac{x_1 + x_2 + \ldots + x_N}{N}.$$

Si los valores del carácter $x_1, x_2, \ldots, x_k$ tienen respectivamente las frecuencias $N_1, N_2, \ldots, N_k$; además, $N_1 + N_2 + \ldots + N_k = N$,

$$\overline{x}_g = \frac{x_1 N_1 + x_2 N_2 + \ldots + x_k N_k}{N},$$

es decir, la media general es la media ponderada de los valores del carácter con pesos iguales a las correspondientes frecuencias.

Nota. Supongamos que un conjunto general de volumen N contiene objetos con distintos valores del carácter X iguales a $x_1, x_2, \ldots, x_N$. Consideremos que de este conjunto extraemos al azar un objeto. La probabilidad de que se extraerá el objeto con valor del carácter, por ejemplo, x_1, evidentemente, es igual a $\frac{1}{N}$. Con igual probabilidad podrá extraerse cualquier otro objeto. Por consiguiente, la magnitud del carácter X se puede considerar como una magnitud aleatoria, cuyos valores posibles $x_1, x_2, \ldots, x_N$ tienen idénticas probabilidades, iguales a $\frac{1}{N}$. Hallemos la esperanza matemática $M(X)$:

$$M(X) = x_1 \cdot \frac{1}{N} + x_2 \cdot \frac{1}{N} + \ldots + x_N \cdot \frac{1}{N} =$$
$$= \frac{x_1 + x_2 + \ldots + x_N}{N} = \bar{x}_g.$$

De este modo, si el carácter investigado X del conjunto general se considera como una magnitud aleatoria, la esperanza matemática del carácter es igual a la media general de este carácter:

$$M(X) = \bar{x}_g.$$

Esta deducción la obtuvimos considerando que todos los objetos del conjunto general tienen distintos valores del carácter. Igual resultado se obtendrá si suponemos que el conjunto general contiene varios objetos con idéntico valor del carácter.

Generalizando el resultado obtenido del conjunto general con distribución continua del carácter X determinamos la media general y, en este caso, como esperanza matemática del carácter:

$$\bar{x}_g = M(X).$$

§ 4. Media muestral

Supongamos que para estudiar el carácter cuantitativo X de un conjunto general se ha extraído la muestra de volumen n.

El valor medio aritmético del carácter del conjunto muestral se llama *media muestral* $\bar{x}_m$.

Si todos los valores $x_1, x_2, \ldots, x_n$ del carácter de la muestra de volumen n son distintos, tendremos que

$$\bar{x}_m = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Si los valores del carácter $x_1, x_2, \ldots, x_k$ tienen respectivamente las frecuencias $n_1, n_2, \ldots, n_k$; además,

$$n_1 + n_2 + \ldots + n_k = n,$$

$$\bar{x}_m = \frac{n_1x_1 + n_2x_2 + \ldots + n_kx_k}{n},$$

o bien

$$\bar{x}_m = \frac{\sum_{i=1}^{k} n_ix_i}{n},$$

es decir, la media muestral es la media ponderada de los valores del carácter de pesos iguales respectivamente a las frecuencias.

Nota. La media muestral hallada por los datos de una muestra es, evidentemente, un número determinado. Si se extraen del mismo conjunto general otras muestras de igual volumen, la media muestral (o valor medio muestral) variará de una muestra a otra. Por consiguiente, la media muestral puede considerarse como una magnitud aleatoria, y, por lo tanto, se puede hablar de distribuciones (teórica y empírica) de la media muestral y de las características numéricas de esta distribución (llamada muestral), en particular, de la esperanza matemática y la dispersión de distribución muestral.

Cabe hacer notar que en los razonamientos teóricos los valores muestrales $x_1, x_2, \ldots, x_n$ del carácter X, obtenidos gracias a observaciones independientes, también se consideran como magnitudes aleatorias $x_1, x_2, \ldots, x_n$ que tienen igual distribución y, por lo tanto, las mismas características numéricas que tienen X.

§ 5. Estimación de la media general según la media muestral. Estabilidad de las medias muestrales

Supongamos que de un conjunto general (como resultado de observaciones independientes sobre el carácter cuantitativo X) se ha extraído una segunda muestra de volumen n con valores del carácter $x_1, x_2, \ldots, x_n$. Sin reducir la generalización de los razonamientos, consideraremos estos valores del carácter distintos. Sea que desconocemos la media general $\bar{x}_g$ y queremos estimarla por los datos de la muestra. Como estimación de la media general tomamos la media muestral

$$\bar{x}_m = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Verifiquemos que $\bar{x}_m$ es la estimación no desviada, es decir, demostremos que la esperanza matemática de esta estimación es igual a $\bar{x}_g$. Vamos a considerar $\bar{x}_m$ como una magnitud aleatoria y $x_1, x_2, \ldots, x_n$, como magnitudes aleatorias independientes igualmente distribuidas de $X_1, X_2, \ldots, X_n$. Puesto que estas magnitudes son igualmente distribuidas, ellas tienen idénticas características numéricas, en particular, igual esperanza matemática que designamos por a. Ya que la esperanza matemática de la media aritmética de las magnitudes aleatorias idénticamente distribuidas es igual a la esperanza matemática de cada una de las magnitudes (cap. VIII, § 9), tendremos que

$$M(\bar{X}_m) = M\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] = a. \qquad (*)$$

Teniendo en cuenta que cada una de las magnitudes $X_1, X_2, \ldots, X_n$ tiene la misma distribución que el conjunto general (que también lo consideramos como una magnitud aleatoria), deducimos que las características numéricas de estas magnitudes y el conjunto general son iguales. En particular, la esperanza matemática a de cada una de las magnitudes es igual a la esperanza matemática del carácter X del conjunto general, es decir,

$$M(X) = \bar{x}_g = a.$$

Sustituyendo en la fórmula (*) la esperanza matemática a por $\bar{x}_g$, finalmente obtenemos

$$M(\bar{X}_m) = \bar{x}_g.$$

Con lo que queda demostrado que la media muestral es la estimación no desviada de la media general.

Se demuestra fácilmente que la media muestral también es la estimación valedera de la media general. En efecto, supongamos que las magnitudes aleatorias $X_1, X_2, \ldots, X_n$ tienen dispersiones limitadas; con derecho aplicamos a estas magnitudes el teorema de Chebishev (caso particular), en virtud del cual al aumentar n la media aritmética de las magnitudes a examinar, es decir, $\overline{X}_m$ tiende respecto de probabilidad a la esperanza matemática a de cada una de las magnitudes o, lo que es igual, a la media general $\overline{x}_g$ (ya que $\overline{x}_g = a$).

De este modo, al aumentar el volumen de la muestra n la media muestral tiende respecto de probabilidad a la media general, lo que significa precisamente que la media muestral es la estimación valedera de la media general.

De lo dicho se deduce que si de varias muestras de volumen suficientemente grande de un mismo conjunto general se hallan las medias muestrales, ellas serán aproximadamente iguales entre sí. En esto radica la propiedad de *estabilidad de las medias muestrales*.

Notemos que si las dispersiones de dos conjuntos son idénticas, la proximidad de las medias muestrales a las generales no depende de la relación entre el volumen de la muestra y el volumen del conjunto general. Ella depende del volumen de la muestra: cuanto mayor es el volumen de la muestra, tanto menos la media muestral se diferencia de la general. Por ejemplo, si de un conjunto se ha escogido el 1% de objetos y de otro, el 4% de objetos, asimismo el volumen de la primera muestra resultó mayor que el de la segunda, tendremos que la primera media muestral se diferenciará menos de la correspondiente media general que la segunda.

Nota. Hemos supuesto que una muestra es repetida. Empero las conclusiones obtenidas son aplicables también para la muestra única si su volumen es bastante menor que el volumen del conjunto general. Esta tesis se utiliza frecuentemente en la práctica.

§ 6. Medias de grupo y general

Admitamos que todos los valores del carácter cuantitativo X del conjunto, indiferentemente general o muestral, están descompuestos en varios grupos. Examinando cada grupo como un conjunto independiente, podemos hallar su media aritmética.

La media aritmética de los valores del carácter, correspondientes al grupo, se llama *media de grupo*.

Ahora conviene introducir un término especial para la media de todo el conjunto.

La media aritmética de los valores del carácter, pertenecientes a todo el conjunto, se llama *media general* $\bar{x}$.

Conociendo las medias de grupo y los volúmenes de los grupos puede hallarse la media general: *la media general es igual a la media aritmética de las medias de grupo, ponderada respecto de los volúmenes de los grupos.*

Omitimos la demostración y damos un ejemplo ilustrativo.

Ejemplo. Hallar la media general del conjunto compuesto de los dos grupos siguientes:

Grupo	primero		segundo	
Valor del carácter	1	6	1	5
Frecuencia	10	15	20	30
Volumen	10 + 15 = 25		20 + 30 + 50.	

SOLUCION. Hallamos las medias de grupo:

$$\bar{x}_1 = \frac{10 \cdot 1 + 15 \cdot 6}{25} = 4;$$

$$\bar{x}_2 = \frac{20 \cdot 1 + 30 \cdot 5}{50} = 3,4.$$

Hallamos la média general según las medias de grupo

$$\bar{x} = \frac{25 \cdot 4 + 50 \cdot 3,4}{25 + 50} = 3,6.$$

Nota. Para simplicar el cálculo de la media general de un conjunto de gran volumen conviene descomponer ésto en varios grupos y hallar las medias de grupo y por ellas la media general.

§ 7. Desviación de la media general y su propiedad

Examinemos el conjunto de los valores, indiferentemente, general o muestral, del carácter cuantitativo X de volumen n:

valor del carácter	x_1	x_2	...	x_k
frecuencia	n_1	n_2	...	n_k;

además $\sum_{i=1}^{k} n_i = n$.

En adelante la suma $\sum_{i=1}^{k}$ la sustituiremos por $\sum_i$.

Hallamos la media general

$$\bar{x} = \frac{\sum n_i x_i}{n}.$$

De aquí

$$\sum n_i x_i = n\bar{x}. \qquad (*)$$

Cabe hacer notar que, siendo $\bar{x}$ una magnitud constante,

$$\sum n_i \bar{x} = \bar{x} \sum n_i = n\bar{x}. \qquad (**)$$

La diferencia $x_i - \bar{x}$ entre el valor del carácter y la media general se llama *desviación*.

Teorema. *La suma de los productos de las desviaciones por las correspondientes frecuencias es igual a cero*

$$\sum n_i (x_i - \bar{x}) = 0.$$

DEMOSTRACION. Teniendo en cuenta (*) y (**) obtenemos

$$\sum n_i (x_i - \bar{x}) = \sum n_i x_i - \sum n_i \bar{x} = n\bar{x} - n\bar{x} = 0.$$

Ejemplo. Dada la distribución del carácter cuantitativo X:

x_i	1	2	3
n_i	10	4	6.

Comprobar que la suma de los productos de las desviaciones por las correspondientes frecuencias es igual a cero.

SOLUCION. Hallamos la media general

$$\bar{x} = \frac{10 \cdot 1 + 4 \cdot 2 + 6 \cdot 3}{20} = 1{,}8.$$

Hallamos la suma de los productos de las desviaciones por las correspondientes frecuencias:

$$\sum n_i (x_i - \bar{x}) = 10 \cdot (1 - 1{,}8) + 4 \cdot (2 - 1{,}8) + \\ + 6 \cdot (3 - 1{,}8) = 8 - 8 = 0.$$

§ 8. Dispersión general

Para determinar la dispersión de los valores del carácter cuantitativo X de un conjunto general alrededor de su valor medio, se introduce una característica general, o sea, la dispersión general.

La media aritmética de los cuadrados de las desviaciones de los valores del carácter del conjunto general respecto de su valor medio $\bar{x}_g$ se llama *dispersión general* D_g.

Si todos los valores $x_1, x_2, \ldots, x_N$ del carácter del conjunto general de volumen N son distintos,

$$D_g = \frac{\sum_{i=1}^{N} (x_i - \bar{x}_g)^2}{N}.$$

Si los valores del carácter $x_1, x_2, \ldots, x_k$ tienen respectivamente las frecuencias $N_1, N_2, \ldots, N_k$; además, $N_1 + N_2 + \ldots + N_k = N$, tendremos que

$$D_g = \frac{\sum_{i=1}^{k} N_i (x_i - \bar{x}_g)^2}{N},$$

es decir, la dispersión general es la media ponderada de los cuadrados de las desviaciones con pesos iguales a las correspondientes frecuencias.

Ejemplo. Un conjunto general está dado por la tabla de distribución

x_i	2	4	5	6
N_i	8	9	10	3.

Hallar la dispersión general.

SOLUCION. Hallamos la media general (§ 3):

$$\bar{x}_g = \frac{8\cdot 2 + 9\cdot 4 + 10\cdot 5 + 3\cdot 6}{8+9+10+3} = \frac{120}{30} = 4.$$

A continuación hallamos la dispersión general:

$$D_g = \frac{8\cdot(2-4)^2 + 9(4-4)^2 + 10(5-4)^2 + 3(6-4)^2}{30} = \frac{54}{30} = 1{,}8.$$

Además de la dispersión, para expresar la dispersión de los valores del carácter de un conjunto general alrededor

de su valor medio se utiliza una característica generla, es decir, la desviación cuadrática media.

La raíz cuadrada de la dispersión general se llama *desviación cuadrática media general* (estándar):

$$\sigma_g = \sqrt{D_g}.$$

§ 9. Dispersión muestral

Para definir la dispresión de los valores observados del carácter cuantitativo de una muestra alrededor de su valor medio $\bar{x}_m$ se introduce la característica general denominada dispersion muestral.

La media aritmética de los cuadrados de las desviaciones de los valores observados del carácter respecto de su valor medio $\bar{x}_m$ se llama *dispersión muestral* D_m.

Si todos los valores $x_1, x_2, \ldots, x_n$ del carácter de la muestra de volumen n son distintos, tendremos que

$$D_m = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_m)^2}{n}.$$

Si los valores del carácter $x_1, x_2, \ldots, x_k$ tienen respectivamente las frecuencias $n_1, n_2, \ldots, n_k$; además, $n_1 + n_2 + \ldots + n_k = n$, entonces

$$D_m = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_m)^2}{n},$$

es decir, la dispersión muestral es la media ponderada de los cuadrados de las desviaciones con pesos iguales a las correspondientes frecuencias.

Ejemplo. Dado un conjunto muestral por la tabla de distriibución

x_i	1	2	3	4
n_i	20	15	10	5,

hallar la dispersión muestral.

SOLUCION. Hallamos la media muestral (§ 4):

$$\bar{x}_m = \frac{20\cdot 1 + 15\cdot 2 + 10\cdot 3 + 5\cdot 4}{20 + 15 + 10 + 5} = \frac{100}{50} = 2.$$

Hallamos la dispersión muestral:

$$D_m = \frac{20(1-2)^2+15(2-2)^2+10(3-2)^2+5(4-2)^2}{50} = \frac{50}{50} = 1.$$

Además de la dispersión, para expesar la dispersión de los valores del carácter del conjunto muestral alrededor de su valor medio se utiliza una característica general, es decir, la desviación cuadrática media.

La raíz cuadrada de la dispersión muestral se llama *desviación cuadrática media muestral* (estándar):

$$\sigma_m = \sqrt{D_m}.$$

§ 10. Fórmula para el cálculo de la dispersión

El cálculo de la dispersión, ya sea muestral o general, puede simplificarse, utilizando el siguiente teorema.

Teorema. *La dispersión es igual a la media de los cuadrados de los valores del carácter menos el cuadrado de la media general*

$$D = \overline{x^2} - [\bar{x}]^2.$$

DEMOSTRACIÓN. El cumplimiento del teorema resulta de las transformaciones

$$D = \frac{\sum n_i (x_i - \bar{x})^2}{n} = \frac{\sum n_i (x_i^2 - 2\bar{x}_i x + [\bar{x}]^2)}{n} =$$

$$= \frac{\sum n_i x_i^2}{n} - 2\bar{x}\,\frac{\sum n_i x_i}{n} + [\bar{x}]^2\,\frac{\sum n_i}{n} = \overline{x^2} - 2\bar{x}\cdot\bar{x} +$$

$$+ [\bar{x}]^2 = \overline{x^2} - [\bar{x}]^2.$$

De este modo

$$D = \overline{x^2} - [\bar{x}]^2,$$

donde

$$\bar{x} = \frac{\sum n_i x_i}{n}, \quad \overline{x^2} = \frac{\sum n_i x_i^2}{n}.$$

Ejemplo. Hallar la dispersión dada la distribución

x_i	1	2	3	4
n_i	20	15	10	5

SOLUCION. Hallamos la media general:

$$\overline{x} = \frac{20\cdot 1+15\cdot 2+10\cdot 3+5\cdot 4}{20+15+10+5} = \frac{100}{50} = 2.$$

Hallamos la media de los cuadrados de los valores del carácter:

$$\overline{x^2} = \frac{20\cdot 1^2+15\cdot 2^2+10\cdot 3^2+5\cdot 4^2}{50} = 5.$$

La dispersión buscada es

$$D = \overline{x^2} - [\overline{x}]^2 = 5 - 2^2 = 1.$$

§ 11. Dispersiones de grupo, dentro de grupo, entre grupos y general

Supongamos que todos los valores del carácter cuantitativo X del conjunto, sea general o muestral, están divididos en k grupos. Examinando cada grupo como un conjunto independiente se puede hallar la media de grupo (§ 6) y la dispersión de los valores del carácter, correspondientes al grupo, respecto a la media de grupo.

Se llama *dispersión de grupo* la dispersión de los valores del carácter, pertenecientes al grupo, respecto de la media de grupo:

$$D_{j\,\text{gru}} = \frac{\sum n_i (x_i - \overline{x}_j)^2}{N_j},$$

donde n_i es la frecuencia del valor x_i;
j, el número del grupo;
$\overline{x}_j$, la media de grupo del grupo j;
$N_j = \sum n_i$, el volumen del grupo j.

Ejemplo 1. Hallar las dispersiones de grupo de un conjunto compuesto de los dos grupos siguientes:

Primer grupo

x_i	n_i
2	1
4	7
5	2

$N_1 = \sum n_i = 10$

Segundo grupo

x_i	n_i
3	2
8	3

$N_2 = \sum n_i = 5.$

SOLUCION. Hallamos las medias de grupo:

$$\bar{x}_1=\frac{\sum n_i x_i}{\sum n_i}=\frac{1\cdot 2+7\cdot 4+2\cdot 5}{10}=4;$$

$$\bar{x}_2=\frac{2\cdot 3+3\cdot 8}{5}=6.$$

Hallamos las dispersiones de grupo buscadas:

$$D_{1\text{gru}}=\frac{\sum n_i(x_i-\bar{x}_1)^2}{N_1}=\frac{1(2-4)^2+7(4-4)^2+2(5-4)^2}{10}=0{,}6;$$

$$D_{2\text{gru}}=\frac{2(3-6)^2+3(8-6)^2}{5}=6.$$

Conocida la dispersión de cada grupo, se puede hallar su media aritmética.

La media aritmética de las dispersiones de grupo, ponderada respecto de los volúmenes de grupos se llama *dispersión dentro de grupo*:

$$D_{\text{den. gru}}=\frac{\sum N_j D_{j\text{gru}}}{n},$$

donde N_j es el volumen del grupo j;

$n=\sum_{j=1}^{k} N_j$ es el volumen de todo el conjunto.

Ejemplo 2. Hallar la dispersión dentro de grupo por los datos del ejemplo 1.

SOLUCION. La dispersión dentro de grupo buscada es igual a

$$D_{\text{den. gru}}=\frac{N_1 D_{1\text{gru}}+N_2 D_{2\text{gru}}}{n}=\frac{10\cdot 0{,}6+5\cdot 6}{15}=\frac{12}{5}.$$

Conocidas las medias de grupo y la media general es posible hallar la dispersión de las medias de grupo con respecto a la media general.

La dispersión de las medias de grupo con respecto a la media general se llama *dispersión entre grupos*:

$$D_{\text{ent. gru}}=\frac{\sum N_j(\bar{x}_j-\bar{x})^2}{n},$$

donde $\overline{x}_j$ es la media de grupo del grupo j;
N_j es el volumen del grupo j;
$\overline{x}$ es la media general;
$n = \sum_{j=1}^{k} N_j$ es el volumen de todo el conjunto.

Ejemplo 3. Hallar la dispersión entre grupos por los datos del ejemplo 1.

SOLUCION. Hallamos la media general

$$\overline{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{1\cdot 2+7\cdot 4+2\cdot 5+2\cdot 3+3\cdot 8}{15} = \frac{14}{3}.$$

Utilizando las magnitudes antes calculadas $\overline{x}_1 = 4$, $\overline{x}_2 = 6$, hallamos la dispersión entre grupos buscada

$$D_{\text{ent. gru}} = \frac{N_1(\overline{x}_1-\overline{x})^2 + N_2(\overline{x}_2-\overline{x})^2}{n} =$$

$$= \frac{10\cdot\left(4-\frac{14}{3}\right)^2 + 5\cdot\left(6-\frac{14}{3}\right)^2}{15} = \frac{8}{9}.$$

A continuación conviene introducir un término especial para la dispersión de todo el conjunto.

La dispersión de los valores del carácter de todo el conjunto respecto de la media general se llama *dispersión general*:

$$D_{\text{gen}} = \frac{\sum n_i (x_i-\overline{x})^2}{n},$$

donde n_i es la frecuencia del valor x_i;
$\overline{x}$ es la media general;
n es el volumen de todo el conjunto.

Ejemplo 4. Hallar la dispersión general por los datos del ejemplo 1.

SOLUCION. Hallamos la dispersión general buscada, teniendo en cuenta que la media general es igual a $\frac{14}{3}$.

$$D_{\text{gen}} = \frac{1\cdot\left(2-\frac{14}{3}\right)^2 + 7\cdot\left(4-\frac{14}{3}\right)^2 + 2\cdot\left(5-\frac{14}{3}\right)^2}{15} +$$

$$+ \frac{2\cdot\left(3-\frac{14}{3}\right)^2 + 3\cdot\left(8-\frac{14}{3}\right)^2}{15} = \frac{148}{45}.$$

Nota. La dispersión general hallada es igual a la suma de las dispersiones dentro de grupo y entre grupos:

$$D_{gen}=\frac{148}{45};$$

$$D_{den.\ gru}+D_{ent.\ gru}=\frac{12}{5}+\frac{8}{9}=\frac{148}{45}.$$

En el párrafo siguiente se demostrará que esta ley se cumple para cualquier conjunto.

§ 12. Suma de dispersiones

Teorema. *Si un conjunto está compuesto de varios grupos, la dispersión general es igual a la suma de las dispersiones de dentro de grupo y de entre grupos:*

$$D_{gen}=D_{den.gru}+D_{ent.gru}.$$

DEMOSTRACION. Para simplificar la demostración suponemos que todo el conjunto de valores del carácter cuantitativo X está dividido en los dos grupos siguientes:

Grupo	primero	segundo
Valor del carácter	x_1 x_2	x_1 x_2
Frecuencia	m_1 m_2	n_1 n_2
Volumen del grupo	$N_1=m_1+m_2$	$N_2=n_1+n_2$
Media de grupo	$\overline{x}_1$	$\overline{x}_2$
Dispersión de grupo	D_{1gru}	D_{2gru}
Volumen de todo el conjunto	$n=N_1+N_2$	

En adelante para comodidad escribiremos en lugar de la suma $\sum_{i=1}^{2}$ solamente el signo $\sum$. Por ejemplo, $\sum m_i=\sum_{i=1}^{2} m_i=$ $=m_1+m_2=N_1$.

También conviene tomar en consideración que si una magnitud constante está afectada por el signo de suma, ésta conviene sacarla fuera de la suma. Por ejemplo, $\sum m_i(\overline{x}_1-\overline{x})^2=(\overline{x}_1-\overline{x})^2\sum m_i=(\overline{x}_1-\overline{x})^2 N_1$.

Hallamos la dispersión general:

$$D_{gen}=\frac{\sum m_i(x_i-\overline{x})^2+\sum n_i(x_i-\overline{x})^2}{n}. \qquad (*)$$

Transformamos el primer sumando del numerador, restando y sumando $\bar{x}_1$:

$$\sum' m_i (x_i - \bar{x})^2 = \sum' m_i [(x_i - \bar{x}_1) + (\bar{x}_1 - \bar{x})]^2 =$$
$$= \sum' m_i (x_i - \bar{x}_1)^2 + 2(\bar{x}_1 - \bar{x}) \sum' m_i (x_i - \bar{x}_1) + \sum' m_i (\bar{x}_1 - \bar{x})^2.$$

Puesto que

$$\sum' m_i = (x_i - \bar{x}_1)^2 = N_1 D_{1gru}$$

(la igualdad se deduce de la correlación $D_{1gru} = \frac{\sum m_i (x_i - \bar{x}_1)^2}{N_i}$ y en virtud de § 7,

$$\sum' m_i (x_i - \bar{x}_1) = 0.$$

el primer sumando toma la forma

$$\sum' m_i (x_i - \bar{x}_1)^2 = N_1 D_{1gru} + N_1 (\bar{x}_1 - \bar{x})^2. \qquad (**)$$

Análogamente se puede representar el segundo sumando del numerador de (*) (restando y sumando $\bar{x}_2$):

$$\sum' n_i (x_i - \bar{x})^2 = N_2 D_{2gru} + N_2 (\bar{x}_2 - \bar{x})^2. \qquad (***)$$

Poniendo (**) y (***) en (*):

$$D_{gen} = \frac{N_1 D_{1gru} + N_2 D_{2gru}}{n} + \frac{N_1 (\bar{x}_1 - \bar{x})^2 + N_2 (\bar{x}_2 - \bar{x})^2}{n} =$$
$$= D_{den.\,gru} + D_{ent.\,gru}.$$

De este modo,

$$D_{gen} = D_{den.\,gru} + D_{ent.\,gru}.$$

En el párrafo precedente se dio un ejemplo que ilustra el teorema demostrado.

Nota. El teorema tiene no sólo valor teórico, sino también un importante valor práctico. Por ejemplo, si gracias a las observaciones se han obtenido varios grupos de valores del carácter, para calcular la dispersión general se puede dejar de unir los grúpos en un conjunto. Por otro lado, si el conjunto tiene gran volumen, conviene dividirlo en varios grupos. En uno u otro caso el cálculo directo de la dispersión general se reemplaza por el cálculo de las dispersiones de los grupos, lo que facilita los cálculos.

§ 13. Estimación de la dispersión general por la dispersión muestral corregida

Supongamos que de un conjunto general como resultado de n observaciones independientes en el carácter cuantitativo X se ha extraído una muestra repetida de volumen n:

valor del carácter	x_1	x_2	...	x_k,
frecuencia	n_1	n_2	...	n_k;

además, $n_1+n_2+\ldots+n_k=n$.

Por la muestra dada hay que estimar (hallar aproximadamente) la dispersión general incógnita D_g. Si como estimación de la dispersión general se toma la dispersión muestral, esta estimación dará lugar a errores sistemáticos, ofreciendo un valor reducido de la dispersión general. Esto se debe a que, como se puede demostrar, la dispersión muestral es una estimación desplazada de la D_g; en otras palabras, la esperanza matemática de la dispersión muestral no es igual a la dispersión general que se estima, sino es igual a

$$M[D_m]=\frac{n-1}{n}D_g.$$

Es fácil «corregir» la dispersión muestral de manera que su esperanza matemática sea igual a la dispersión general. Para ello es suficiente multiplicar D_m por la fracción $\frac{n}{n-1}$. Con esto obtenemos la dispersión corregida que de ordinario se designa por s^2:

$$s^2=\frac{n}{n-1}D_m=\frac{n}{n-1}\frac{\sum_{i=1}^{k}n_i(x_i-\bar{x}_m)^2}{n}=\frac{\sum_{i=1}^{k}n_i(x_i-\bar{x}_m)^2}{n-1}.$$

La dispersión corregida es, evidentemente, la estimación no desviada de la dispersión general. En efecto,

$$M[s^2]=M\left[\frac{n}{n-1}D_m\right]=\frac{n}{n-1}M[D_m]=\frac{n}{n-1}\cdot\frac{n-1}{n}D_g=D_g.$$

De este modo, como estimación de la dispersión general se admite la dispersión corregida

$$s^2=\frac{\sum_{i=1}^{k}n_i(x_i-\bar{x}_m)^2}{n-1}.$$

Para estimar la desviación cuadrática media de un conjunto general se utiliza la desviación cuadrática media «corregida» que es igual a la raíz cuadrada de la dispersión corregida:

$$s = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_m)^2}{n-1}}.$$

Subrayemos que s no es la estimación no desviada; para expresar este hecho hemos escrito y escribiremos en adelante así: desviación cuadrática media «corregida».

Nota. Comparando las fórmulas

$$D_m = \frac{\sum n_i (x_i - \bar{x}_m)^2}{n} \quad \text{y} \quad s^2 = \frac{\sum n_i (x_i - \bar{x}_m)^2}{n-1}$$

vemos que ellas se diferencian sólo por los denominadores. Evidentemente, para valores suficientemente grandes n del volumen de la muestra, las dispersiones muestral y corregida se diferencian poco. En la práctica se utiliza la dispersión corregida, si aproximadamente $n < 30$.

§ 14. Exactitud de estimación, probabilidad fiducial (fiabilidad). Intervalo confidencial

La estimación se llama *puntual* cuando ésta se determina *por un número*. Todas las estimaciones estudiadas antes son puntuales. Para una muestra de pequeño volumen la estimación puntual puede diferenciarse bastante del parámetro que se estima, es decir, da lugar a errores groseros. Por esa causa, para un volumen pequeño de la muestra hay que utilizar las estimaciones de intervalo.

La estimación determinada *por dos números*, es decir, los extremos de un intervalo, se llama *de intervalo*. Las estimaciones de intervalo permiten establecer la precisión y la fiabilidad de las estimaciones (el sentido de estos conceptos se aclara más adelante).

Supongamos que la característica estadística Θ^* hallada por los datos de la muestra sirve de estimación del parámetro desconocido Θ. Vamos a considerar Θ como un número constante (Θ puede ser también una magnitud aleatoria). Está claro que Θ^* determina con tanta mayor precisión el parámetro Θ, cuanto menor es la magnitud absoluta de la diferencia $|\Theta - \Theta^*|$. En otras palabras, si $\delta > 0$ y

$|\Theta - \Theta^*| < \delta$, cuanto menor es δ, tanto más exacta es la estimación. Por consiguiente, el número positivo δ caracteriza la *precisión de la estimación*.

Sin embargo, los métodos estadísticos no permiten afirmar categóricamente que la estimación Θ^* satisface la desigualdad $|\Theta - \Theta^*| < \delta$; sólo es posible hablar de la probabilidad γ, con la cual se cumple esta desigualdad.

La probabilidad γ con la que se cumple la desigualdad $|\Theta - \Theta^*| < \delta$, se llama *fiabilidad* (*probabilidad fiducial*) de la estimación Θ según Θ^*. Generalmente la fiabilidad de estimación se da previamente; además, como γ se toma un número proximo a la unidad. Con más frecuencia se prefija la fiabilidad igual a 0,95; 0,99 y 0,999.

Supongamos que la probabilidad de que $|\Theta - \Theta^*| < \delta$ sea igual a γ:

$$P\,[\,|\Theta - \Theta^*| < \delta\,] = \gamma.$$

Sustituyendo la desigualdad $|\Theta - \Theta^*| < \delta$ por la doble desigualdad equivalente $-\delta < \Theta - \Theta^* < \delta$, o bien

$$\Theta^* - \delta < \Theta < \Theta^* + \delta,$$

tendremos

$$P\,[\Theta^* - \delta < \Theta < \Theta^* + \delta] = \gamma.$$

Esta correlación se interpreta así: la probabilidad de que el intervalo $(\Theta^* - \delta,\ \Theta^* + \delta)$ incluye en sí (recubre) el parámetro incógnita Θ, es igual a γ.

El intervalo $(\Theta^* - \delta,\ \Theta^* + \delta)$ que recubre el parámetro desconocido con la fiabilidad γ prefijada se llama *intervalo de confianza o confidencial*.

Nota. El intervalo $(\Theta^* - \delta,\ \Theta^* + \delta)$ tiene extremos aleatorios (llamados limites de confianza o confidenciales). En efecto, en distintas muestras se obtienen distintos valores de Θ. Por lo tanto, de una muestra a otra variarán también los extremos del intervalo confidencial, es decir, los limites confidenciales son propiamente magnitudes aleatorias, o sea, funciones de $x_1, x_2, \ldots, x_n$.

Ya que la magnitud aleatoria es el intervalo confidencial y no el parámetro a estimar Θ, sería más correcto hablar sobre la probabilidad de que el intervalo de confianza recubra Θ y no sobre la probabilidad de que Θ caiga en el intervalo de confianza.

El método de los intervalos de confianza o confidencial ha sido elaborado por el estadístico norteamericano Y. Neyman, partiendo de las ideas del estadístico inglés R. Fisher.

§ 15. Intervalos de confianza para estimar la esperanza matemática de distribución normal cuando se conoce σ

Supongamos que el carácter cuantitativo X de un conjunto general está distribuido normalmente; además, se conoce la desviación cuadrática media σ de esta distribución. Se necesita estimar la esperanza matemática desconocida a por la media muestral $\bar{x}$. Tratemos de hallar los intervalos confidenciales que recubren el parámetro a con fiabilidad γ.

Consideremos la media muestral $\bar{x}$ como una magnitud aleatoria $\bar{X}$ ($\bar{x}$ varía de una muestra a otra) y los valores muestrales del carácter $x_1, x_2, \ldots, x_n$, como magnitudes aleatorias independientes igualmente distribuidas $X_1, X_2, \ldots, X_n$ (estos números también varían de una muestra a otra). En otras palabras, la esperanza matemática de cada una de estas magnitudes es igual a a y la desviación cuadrática media es σ.

Admitiremos sin demostración que si la magnitud aleatoria X está distribuida normalmente, la media muestral $\bar{X}$ hallada por observaciones independientes también está distribuida normalmente. Los parámetros de la distribución de $\bar{X}$ son siguientes (cap. VIII, § 9):

$$M(\bar{X}) = a, \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

Necesitamos que se cumpla la correlación

$$P(|\bar{X} - a| < \delta) = \gamma,$$

donde γ es la fiabilidad prefijada.

Utilizando la fórmula (cap. XII, § 6)

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right),$$

sustituyendo X por $\bar{X}$ y σ por $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, obtenemos

$$P(|\bar{X} - a| < \delta) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi(t),$$

donde $t = \frac{\delta\sqrt{n}}{\sigma}$.

Hallando de la última igualdad $\delta = t\frac{\sigma}{\sqrt{n}}$, podemos escribir

$$P\left(|\overline{X}-a|<t\frac{\sigma}{\sqrt{n}}\right)=2\Phi(t).$$

A partir de que la probabilidad P está prefijada y es igual a γ, finalmente tendremos (para obtener una fórmula de trabajo, la media muestral la designamos nuevamente por $\overline{x}$):

$$P\left(\overline{x}-t\frac{\sigma}{\sqrt{n}}<a<\overline{x}+t\frac{\sigma}{\sqrt{n}}\right)=2\Phi(t)=\gamma.$$

La correlación obtenida nos dice que: con la fiabilidad γ se puede afirmar que el intervalo confidencial $\left(\overline{x}-t\frac{\sigma}{\sqrt{n}}, \overline{x}+t\frac{\sigma}{\sqrt{n}}\right)$ recubre el parámetro incógnita a; la exactitud de estimación $\delta = t\cdot\frac{\sigma}{\sqrt{n}}$.

De este modo, el problema planteado antes está completamente resuelto.

Cabe agregar que el número t se determina de la igualdad $2\Phi(t)=\gamma$, o bien $\Phi(t)=\frac{\gamma}{2}$; por la tabla de la función de Laplace (suplemento 2) se halla el argumento t, al que corresponde un valor de la función de Laplace igual a $\frac{\gamma}{2}$.

Nota. 1. La estimación $|\overline{x}-a|<t\cdot\frac{\sigma}{\sqrt{n}}$ se llama clásica. De la fórmula $\delta=t\cdot\frac{\sigma}{\sqrt{n}}$ que determina la exactitud de la estimación clásica se pueden hacer las deducciones siguientes:

1) al aumentar el volumen de la muestra n el número δ decrece y, por lo tanto, se eleva la exactitud de la estimación;

2) el incremento de la fiabilidad de la estimación $\gamma=2\Phi(t)$ da lugar al aumento de t ($\Phi(t)$ es una función creciente) y, por consiguiente, también al aumento de δ; en otras palabras, el incremento de la fiabilidad de estimación clásica conduce a la disminución de su exactitud.

Ejemplo. La magnitud aleatoria X tiene distribución normal con desviación cuadrática media conocida $\sigma = 3$.

Hallar los intervalos confidenciales para estimar la esperanza matemática incógnita a según la media muestral $\bar{x}$, si el volumen de la muestra $n = 36$ y la fiabilidad de estimación es $\gamma = 0,95$.

SOLUCION. Hallamos t. De la correlación $2\Phi(t) = 0,95$ obtenemos $\Phi(t) = 0,475$. Por la tabla (suplemento 2) hallamos

$$t = 1,96.$$

Hallamos la exactitud de estimación:

$$\delta = \frac{t \cdot \sigma}{\sqrt{n}} = \frac{1,96 \cdot 3}{\sqrt{36}} = 0.98.$$

Los intervalos confidenciales son:

$$(\bar{x} - 0,98;\ \bar{x} + 0,98).$$

Por ejemplo, si $x = 4,1$, el intervalo confidencial tiene los siguientes límites de confianza:

$$\bar{x} - 0,98 = 4,1 - 0,98 = 3,12;$$
$$\bar{x} + 0,98 = 4,1 + 0,98 = 5,08.$$

Por consecuencia, los valores del parámetro incógnito a, que concuerdan con los datos de la muestra, satisfacen la desigualdad

$$3,12 < a < 5,08.$$

Cabe hacer notar que sería erróneo escribir

$$P(3,12 < a < 5,08) = 0,95$$

En efecto, dado que a es una magnitud constante, o bien se encuentra en el intervalo hallado (en tal caso el suceso $3,12 < a < 5,08$ es cierto y su probabilidad es igual a la unidad), o bien no se encuentra en él (en ese caso el suceso $3,12 < a < 5,08$ es imposible y su probabilidad es igual a cero). En otras palabras, la probabilidad fiducial no hay que vincularla con el parámetro que se estima; ella está ligada solamente con los límites del intervalo confidencial, los que, como se indicó, varían de una muestra a otra.

Aclaremos el significado de la fiabilidad prefijada. La fiabilidad $\gamma = 0,95$ indica que si se ha realizado un número

suficientemente grande de muestras, el 95% de ellas determina tales intervalos confidenciales, en los cuales el parámetro está realmente contenido; sólo en el 5% de los casos puede salir de los límites del intervalo de confianza.

Nota 2. Si hay que estimar la esperanza matemática y se conocen previamente la exactitud δ y la fiabilidad γ, el volumen mínimo de la muestra que garantiza esta exactitud se halla por la fórmula

$$n=\frac{t^2\sigma^2}{\delta^2}$$

$\left(\text{resultado de la igualdad } \delta = t\frac{\sigma}{\sqrt{n}}\right)$.

§ 16. Intervalos de confianza para estimar la esperanza matemática de distribución normal para σ incógnita

Supongamos que el carácter cuantitativo X de un conjunto general está distribuido normalmente; además, la desviación cuadrática media σ es incógnita. Hay que estimar la esperanza matemática desconocida a mediante los intervalos de confianza. Desde luego, es imposible utilizar los resultados del párrafo anterior, en el que σ se supone conocida.

Resulta que por los datos de la muestra se puede formar la magnitud aleatoria (sus valores posibles los designaremos por t),

$$T=\frac{\overline{X}-a}{\frac{S}{\sqrt{n}}},$$

que tiene la distribución t de Student con $k = n - 1$ grados de libertad (véase la explicación al final del párrafo); aquí $\overline{X}$ es la media muestral, S es la desviación cuadrática media «corregida», n es el volumen de la muestra.

La función diferencial de la distribución t de Student

$$S(t, n)=B_n\left[1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}},$$

donde $B_n=\dfrac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)}$.

Como podemos apreciar la distribución t de Student se determina por el parámetro n, o sea, el volumen de la muestra (o lo que es igual, por el número de grados de libertad $k = n - 1$) y no depende de los parámetros incógnitos a y σ; esta particularidad es su gran mérito). Puesto que $S(t, n)$ es una función par de t, la probabilidad de que se cumpla la desigualdad $\left|\frac{\overline{X}-a}{\frac{S}{\sqrt{n}}}\right| < \gamma$ se determina así (cap. XI, § 2, nota):

$$P\left(\left|\frac{\overline{X}-a}{\frac{S}{\sqrt{n}}}\right| < t_\gamma\right) = 2\int_0^{t_\gamma} S(t, n)\,dt = \gamma.$$

Sustituyendo la desigualdad entre paréntesis por la doble desigualdad equivalente, obtenemos

$$P\left(\overline{X} - t_\gamma \frac{S}{\sqrt{n}} < a < \overline{X} + t_\gamma \frac{S}{\sqrt{n}}\right) = \gamma.$$

De este modo, mediante la distribución t de Student hemos hallado el intervalo confidencial $\left(\overline{x} - t_\gamma \frac{s}{\sqrt{n}},\ \overline{x} + t_\gamma \frac{s}{\sqrt{n}}\right)$ que recubre el parámetro incógnito a con la fiabilidad γ. Aquí las magnitudes aleatorias $\overline{X}$ y S están reemplazadas por las magnitudes no aleatorias $\overline{x}$ y s halladas por la muestra. Por la tabla (suplemento 3), por n y γ dados se puede hallar t_γ.

Ejemplo. El carácter cuantitativo X de un conjunto general está distribuido normalmente. Por la muestra de volumen $n = 16$ se halló la media muestral $\overline{x} = 20{,}2$ y la desviación cuadrática media «corregida» $s = 0{,}8$. Estimar la esperanza matemática desconocida mediante el intervalo confidencial con la fiabilidad 0,95.

SOLUCION. Hallamos t_γ. Utilizando la tabla (suplemento 3) por $\gamma = 0{,}95$ y $n = 16$, obtendremos $t_\gamma = 2{,}13$.

Encontramos los límites confidenciales:

$$\overline{x} - t_\gamma \frac{s}{\sqrt{n}} = 20{,}2 - 2{,}13 \frac{0{,}8}{\sqrt{16}} = 19{,}774,$$

$$\overline{x} + t_\gamma \frac{s}{\sqrt{n}} = 20{,}2 + 2{,}13 \frac{0{,}8}{\sqrt{16}} = 20{,}626.$$

Así pues, con la fiabilidad 0,95 el parámetro incógnita a está contenido en el intervalo de confianza $19{,}774 < a <$ $< 20{,}626$.

Nota. De las correlaciones límites

$$\lim_{n\to\infty} B_n = \frac{1}{\sqrt{2\pi}}, \qquad \lim_{n\to\infty}\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}} = e^{-\frac{t^2}{2}}$$

se deduce que al crecer ilimitadamente el volumen de la muestra n la distribución t de Student tiende a la normal. Por eso, cuando $n > 30$ se puede utilizar la distribución normal en lugar de la distribución t de Student.

Sin embargo, es importante señalar que *para pequeñas muestras* ($n < 30$), en especial para pequeños valores de n, el reemplazo de la distribución por la normal da lugar a errores groseros, precisamente, a un estrechamiento injustificado del intervalo confidencial, es decir, al aumento de la exactitud de estimación. Por ejemplo, si $n = 5$ y $\gamma = 0{,}99$, utilizando la distribución t de Student hallamos $t_\gamma = 4{,}6$, en tanto que utilizando la función de Laplace hallamos $t_\gamma =$ $= 2{,}58$, o sea que el intervalo confidencial en el último caso resulta más estrecho que el hallado por la distribución t de Student.

El hecho de que la distribución t de Student da resultados no del todo determinados para una muestra pequeña (intervalo confidencial amplio), en general no significa que el método t de Student sea inseguro; esto se debe a que una pequeña muestra, naturalmente, contiene poca información sobre el carácter que nos interesa.

Explicación. Antes se dijo que (cap. XII, § 14) si Z es una magnitud normal cuando $M(Z) = 0$, $\sigma(Z) = 1$, y V es una magnitud independiente de Z distribuida por la ley de χ^2 con k grados de libertad, la magnitud

$$T = \frac{Z}{\sqrt{\frac{V}{k}}} \qquad (*)$$

está distribuida por la ley t de Student con k grados de libertad.

Supongamos que el carácter cuantitativo X de un conjunto general está distribuido normalmente, siendo $M(X) = a$, $\sigma(X) = \sigma$. Si de este conjunto se extrae una muestra de

volumen n y por ella se hallan las medias muestrales, se puede demostrar que la media muestral está distribuida normalmente; asimismo (cap. VIII, § 9)

$$M(\overline{X}_m)=a, \qquad \sigma(\overline{X}_m)=\frac{\sigma}{\sqrt{n}}.$$

En tal caso, la magnitud aleatoria

$$Z=\frac{\overline{X}_m-a}{\frac{\sigma}{\sqrt{n}}} \qquad (**)$$

también tiene distribución normal, como una función lineal de argumento normal $\overline{X}_m$ (cap. XII, § 10, nota); al mismo tiempo $M(Z)=0$, $\sigma(Z)=1$.

Se ha demostrado que la magnitud aleatoria

$$V=\frac{(n-1)S^2}{\sigma^2} \qquad (***)$$

independiente de Z (S^2 es la dispersión muestral corregida) está distribuida por la ley de χ^2 con $k=n-1$ grados de libertad.

Por lo tanto, poniendo (**) y (***) en (*) obtenemos la magnitud

$$T=\frac{(\overline{X}_m-a)\sqrt{n}}{S},$$

que está distribuida por la ley t de Student con $k=n-1$ grados de libertad.

§ 17. Estimación del valor real de la magnitud a medir

Supongamos que se realizan n mediciones independientes de igual precisión de cierta magnitud física, cuyo valor real a no se conoce. Los resultados de las mediciones individuales los consideraremos como magnitudes aleatorias $X_1, X_2, \ldots, X_n$. Estas magnitudes son independientes (mediciones independientes), tienen la misma esperanza matemática a (valor real de la magnitud que se mide), idénticas dispersiones σ^2 (mediciones de igual precisión) y están distribuidas normalmente (esta suposición es confirmada por la experiencia). Por consiguiente, todas las hipótesis hechas al deducir los intervalos confidenciales en los dos párrafos precedentes se cumplen y, por lo tanto,

tenemos razón de utilizar las fórmulas obtenidas en ellas. En otras palabras, el valor real de la magnitud a medir se puede estimar por la media aritmética de los resultados de las mediciones individuales mediante los intervalos de confianza. Ya que generalmente σ es una incógnita hay que utilizar las fórmulas del § 16.

Ejemplo. Por los datos de nueve mediciones independientes de igual precisión de una magnitud física se ha hallado la media aritmética de los resultados de las mediciones individuales $\bar{x} = 42,319$ y la desviación cuadrática media «corregida» $s = 5,0$. Hay que estimar el valor real de la magnitud que se mide con fiabilidad $\gamma = 0,95$.

SOLUCION. El valor real de la magnitud que se mide es igual a su esperanza matemática. Por lo tanto el problema se reduce a estimar la esperanza matemática (para σ incógnita) mediante el intervalo confidencial

$$\bar{x} - t_\gamma \frac{s}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{s}{\sqrt{n}},$$

que recubre a con la fiabilidad prefijada $\gamma = 0,95$.

Utilizando la tabla (suplemento 3) por $\gamma = 0,95$ y $n = 9$, hallamos $t_\gamma = 2,31$.

Hallamos la exactitud de estimación:

$$t_\gamma \cdot \frac{s}{\sqrt{n}} = 2,31 \cdot \frac{5}{3} = 3,85.$$

Hallamos los límites de confianza:

$$\bar{x} - t_\gamma \frac{s}{\sqrt{n}} = 42,319 - 3,85 = 38,469;$$

$$\bar{x} + t_\gamma \frac{s}{\sqrt{n}} = 42,319 + 3,85 = 46,169.$$

De este modo, con fiabilidad de 0,95 el valor real de la magnitud que se mide está dentro del intervalo de confianza

$$38,469 < a < 46,169.$$

§ 18. Intervalos de confianza para estimar la desviación cuadrática media σ de una distribución normal

Supongamos que el carácter cuantitativo X de un conjunto general está distribuido normalmente. Se necesita estimar la desviación cuadrática media general desconocida σ por la

desviación cuadrática media muestral «corregida» s. Tratemos de hallar los intervalos confidenciales que recubren el parámetro σ con la fiabilidad dada γ.

Necesitamos que se cumple la correlación

$$P(|\sigma - s| < \delta) = \gamma,$$

o bien

$$P(s - \delta < \sigma < s + \delta) = \gamma.$$

Para poder utilizar la tabla transformamos la doble desigualdad

$$s - \delta < \sigma < s + \delta$$

en la desigualdad equivalente

$$s\left(1 - \frac{\delta}{s}\right) < \sigma < s\left(1 + \frac{\delta}{s}\right).$$

Poniendo $\frac{\delta}{s} = q$, obtenemos

$$s(1 - q) < \sigma < s(1 + q). \qquad (*)$$

Queda hallar q. Para ello introducimos en el examen la magnitud aleatoria «ji»:

$$\chi = \frac{S}{\sigma}\sqrt{n-1},$$

donde n es el volumen de la muestra.

Como se indicó [§ 16, explicación, correlación (***)] la magnitud $\frac{S^2(n-1)}{\sigma^2}$ está distribuida por la ley de χ^2, por esto su raíz cuadrada se designa por χ.

La función diferencial de la distribución χ tiene la forma (véase explicación al final del párrafo)

$$R(\chi, n) = \frac{\chi^{n-2} e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}. \qquad (**)$$

Como vemos, esta distribución no depende del parámetro que se estima σ, sino depende solamente del volumen de la muestra n.

Transformemos la desigualdad (*) de tal modo que tome la forma

$$\chi_1 < \chi \leqslant \chi_2.$$

La probabilidad de que esta desigualdad (cap. XI, § 2) sea igual a la probabilidad prefijada γ, es decir

$$\int_{\chi_1}^{\chi_2} R(\chi, n)\, d\chi = \gamma.$$

Suponiando que $q < 1$, la desigualdad (*) la escribimos así:

$$\frac{1}{S(1+q)} < \frac{1}{\sigma} < \frac{1}{S(1-q)}.$$

Multiplicando todos miembros de la desigualdad por $S\sqrt{n-1}$, obtenemos

$$\frac{\sqrt{n-1}}{1+q} < \frac{S\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q},$$

o bien

$$\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}.$$

La probabilidad de que esta desigualdad y, por lo tanto, también la desigualdad (*) equivalente a ella se cumpla, es igual a

$$\int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} R(\chi, n)\, d\chi = \gamma.$$

De esta ecuación se puede hallar q por n y γ prefijados. Prácticamente, para hallar q se utiliza la tabla (suplemento 4).

Calculando por la muestra s y hallando por la tabla q obtenemos el intervalo confidencial buscado (*) que recubre σ con la probabilidad dada γ, es decir, el intervalo

$$s(1-q) < \sigma < s(1+q).$$

Ejemplo 1. El carácter cuantitativo X de un conjunto general está distribuido normalmente. Por la muestra de volumen $n = 25$ se halla la desviación cuadrática media «corregida» $s = 0{,}8$. Hallar el intervalo de confianza que recubre la desviación cuadrática media general σ con la fiabilidad 0,95.

SOLUCION. Por la tabla (suplemento 4) según los datos $\gamma = 0{,}95$ y $n = 25$ hallamos $q = 0{,}32$.

El intervalo de confianza buscado (*) es:

$$0{,}8\ (1 - 0{,}32) < \sigma < 0{,}8\ (1 + 0{,}32),$$

o bien

$$0{,}544 < \sigma < 1{,}056.$$

Nota. Antes se supuso que $q < 1$. Si $q > 1$, la desigualdad (*) toma la forma (teniendo en cuenta que $\sigma > 0$)

$$0 < \sigma < s\,(1 + q),$$

o bien (después de transformaciones análogas al caso $q < 1$)

$$\frac{\sqrt{n-1}}{1+q} < \chi < \infty.$$

Por lo tanto, los valores de $q > 1$ pueden ser hallados de la ecuación

$$\int\limits_{\frac{\sqrt{n-1}}{1+q}}^{\infty} R(\chi, n)\, d\chi = \gamma.$$

Prácticamente para hallar los valores de $q > 1$, correspondientes a distintos n y γ prefijados, se utiliza la tabla (suplemento 4).

Ejemplo 2. El carácter cuantitativo X de un conjunto general está distribuido normalmente. Por la muestra de volumen $n = 10$ se halló la desviación cuadrática media «corregida» $s = 0{,}16$. Hallar el intervalo confidencial que recubre la desviación cuadrática media general σ con la fiabilidad 0,999.

SOLUCION. Por la tabla (sumplemento 4) según los $\gamma = 0{,}999$ y $n = 10$ dados hallamos $q = 1{,}80$ $(q > 1)$. El intervalo de confianza buscado es:

$$0 < \sigma < 0{,}16\ (1 + 1{,}80)$$

o bien

$$0 < \sigma < 0{,}448.$$

Explicación. Demostramos que la función diferencial de distribución χ tiene la forma (**).

Si la magnitud aleatoria X está distribuida por la ley de χ^2 con $k = n - 1$ grados de libertad, su función diferen-

cial es (cap. XII, § 13)

$$f(x)=\frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)},$$

o bien, después de sustituir $k=n-1$,

$$f(x)=\frac{x^{\frac{n-3}{2}}e^{-\frac{x}{2}}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}.$$

Utilizamos la fórmula (cap. XII, § 10)

$$g(y)=f[\psi(y)]\cdot|\psi'(y)|,$$

para hallar la distribución de la función $\chi=\varphi(X)=\sqrt{X}$ $(\chi>0)$. De aquí la función inversa

$$x=\psi(\chi)=\chi^2$$

y

$$\psi'(\chi)=2\chi.$$

Puesto que

$$\chi>0, \text{ tendremos que } |\psi'(\chi)|=2\chi.$$

Por lo tanto,

$$g(\chi)=f[\psi(\chi)]\cdot|\psi'(\chi)|=\frac{(\chi^2)^{\frac{n-3}{2}}e^{-\frac{\chi^2}{2}}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}2\chi.$$

Cumpliendo transformaciones elementales y sustituyendo $g(\chi)$ por $R(\chi, n)$, finalmente obtenemos

$$R(\chi, n)=\frac{\chi^{n-2}e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)}.$$

§ 19. Estimación de la exactitud de mediciones

En la teoría de los errores la precisión de las mediciones (precisión del instrumento) se admite en caracterizar mediante la desviación cuadrática media σ de los errores aleatorios

de las mediciones. Para estimar σ se utiliza la desviación cuadrática media «corregida» s.

Puesto que, generalmente, los resultados de las mediciones son mutuamente independientes, tienen igual esperanza matemática (valor real de la magnitud que se mide) e idéntica dispersión (en el caso de mediciones de igual precisión), la teoría expuesta en el párrafo anterior es aplicable para la estimación de las mediciones.

Ejemplo. Por 15 mediciones de igual precisión se halló la desviación cuadrática media «corregida» $s = 0,12$. Hallar la exactitud de las mediciones con la fiabilidad 0,99.

SOLUCIÓN. La exactitud de las mediciones se caracteriza por la desviación cuadrática media σ de los errores aleatorios, por lo cual el problema se reduce a hallar el intervalo de confianza (*) (§ 18) que recubre σ con la fiabilidad 0,99 dada.

Por la tabla (suplemento 4) según $\gamma = 0,99$ y $n = 15$ hallamos $q = 0.73$. El intervalo confidencial buscado es:

$$0,12\ (1 - 0,73) < \sigma < 0,12\ (1 + 0,73),$$

o bien

$$0,03 < \sigma < 0,21.$$

§ 20. Otras características de la serie de variación

Además de la media muestral y la dispersión muestral se utilizan también otras características de la serie de variación. Citamos las más importantes de ellas.

La variante que tiene frecuencia máxima se llama *moda* (M_0). Por ejemplo, para la serie

variante	1	4	7	9
frecuencia	5	1	20	6

la moda es igual a 7.

La variante que divide la serie de variación en dos partes, iguales en númeró de variantes, se llama *mediana* m_e. Si el número de variantes es impar, es decir $n = 2k + 1$, tendremos que $m_e = x_{k+1}$; cuando es par $n = 2k$ la mediana es

$$m_e = \frac{x_k + x_{k+1}}{2}.$$

Por ejemplo, para la serie

$$2\ 3\ 5\ 6\ 7$$

la mediana es igual a 5; para la serie

$$2\ 3\ 5\ 6\ 7\ 9$$

la mediana es igual a $\frac{5+6}{2} = 5{,}5$.

La diferencia entre las variaciones máxima y mínima se llama *amplitud de variación R*:

$$R = x_{\text{máx}} - x_{\text{mín}}.$$

Por ejemplo, para la serie

$$1\ 3\ 4\ 5\ 6\ 10$$

la amplitud es igual a $10 - 1 = 9$.

La amplitud es una característica elemental de dispersión de la serie de variación.

La media aritmética de las desviaciones absolutas se llama *desviación absoluta media* Θ:

$$\Theta = \frac{\sum n_i |x_i - \bar{x}_\text{m}|}{\sum n_i}.$$

Por ejemplo, para la serie

x_i	1	3	6	16
n_i	4	10	5	1

tenemos

$$\bar{x}_\text{m} = \frac{4\cdot 1 + 10\cdot 3 + 5\cdot 6 + 1\cdot 16}{4+10+5+1} = \frac{80}{20} = 4;$$

$$\Theta = \frac{4\cdot|1-4| + 10\cdot|3-4| + 5\cdot|6-4| + 1\cdot|16-4|}{20} = 2{,}2.$$

La desviación absoluta media sirve para la característica de dispersión de la serie de variación.

La relación entre la desviación cuadrática media muestral y la media muestral, expresada en tantos por ciento, se llama *coeficiente de variación V*:

$$V = \frac{\sigma_\text{m}}{\bar{x}_\text{m}} \cdot 100\ \%.$$

El coeficiente de variación sirve para comparar las magnitudes de las dispersiones de dos series de variación: aquella de las series cuyo coeficiente de variación es mayor, tiene mayor dispersión.

Nota. Antes se supuso que la serie de variación está formada por los datos de la muestra, por eso todas las características descriptas se llaman *muestrales*; si la serie de variación está formada por los datos de un conjunto general, las características se llaman *generales*.

Problemas

1. Hallar las medias de grupo de un conjunto compuesto de dos grupos:

primer grupo	x_i	0,1	0,4	0,6
	n_i	3	2	5;
segundo grupo	x_i	0,1	0,3	0,4
	n_i	10	4	6

Respuesta $\bar{x}_1 = 0,41$; $\bar{x}_2 = 0,23$.

2. Hallar la media general según los datos del primer problema por dos métodos: a) uniendo ambos grupos en un conjunto; b) utilizando las medias de grupo halladas en el primer problema.

Respuesta $\bar{x} = 0,29$.

3. Dada la distribución de un conjunto estadístico

x_i	1	4	5
n_i	6	11	3,

verificar que la suma de los productos de las desviaciones por las correspondientes frecuencias es igual a cero.

4. Dada la distribución de un conjunto estadístico

x_i	4	7	10	15
n_i	10	15	20	5,

hallar la dispersión del conjunto: a) partiendo de la definición de dispersión; b) utilizando la fórmula $D = \overline{x^2} - [\bar{x}]^2$.

Respuesta $D = 9,84$.

5. Hallar las dispersiones dentro de grupo, entre grupos y general de un conjunto compuesto de tres grupos:

primer grupo	x_i	1	2	8
	n_i	30	15	5;

$$\begin{cases} \text{segundo grupo} & x_i & 1 & 6 \\ & n_i & 10 & 15; \\ \text{tercer grupo} & x_i & 3 & 8 \\ & n_i & 20 & 5. \end{cases}$$

Respuesta $D_{\text{den. gru}} = 4{,}6$; $D_{\text{ent. gru}} = 1$; $D_{\text{gen}} = 5{,}6$.

6. Hallar las dispersiones dentro de grupo, entre grupos y general de un conjunto compuesto de dos grupos:

$$\begin{matrix} \text{primer grupo} & x_i & 2 & 7 \\ & n_i & 6 & 4; \\ \text{segundo grupo} & x_i & 2 & 7 \\ & n_i & 2 & 8. \end{matrix}$$

Respuesta $D_{\text{den. gru}} = 5$; $D_{\text{ent. gru}} = 1$; $D_{\text{gen}} = 6$.

7. Hallar las dispersiones muestral y corregida de una serie de variación compuesta por los datos de las muestras

variante	1	2	5	8	9
frecuencia	3	4	6	4	3.

Respuesta $\sigma_m^2 = 8{,}4$; $s^2 = 8{,}84$.

En los problemas 8 y 9 se dan la desviación cuadrática media, la media muestral y el volumen de la muestra del carácter distribuido normalmente. Hallar los intervalos de confianza para estimar la esperanza matemática incógnita con la fiabilidad dada.

8. $\sigma = 2$, $\bar{x}_m = 5{,}40$, $n = 10$, $\gamma = 0{,}95$.

Respuesta $4{,}16 < a < 6{,}64$.

9. $\sigma = 3$, $\bar{x}_m = 20{,}12$, $n = 25$, $\gamma = 0{,}99$.

Respuesta $18{,}57 < a < 21{,}67$.

10. Hallar el volumen mínimo de la muestra, para el cual con la fiabilidad 0,95 la exactitud de estimación de la esperanza matemática del carácter distribuido normalmente según la media muestral será igual a 0,2, si la desviación cuadrática media es igual a 2.

Observación. Véase la nota 2, § 15.

Respuesta $n = 385$.

En los problemas 11 y 12 se dan la desviación cuadrática media «corregida», la media muestral y el volumen de una pequeña muestra del carácter distribuido normalmente. Hallar mediante la distribución de Student los intervalos confidenciales para estimar la esperanza matemática incógnita con la fiabilidad dada.

11. $s = 1{,}5,\ \overline{x}_m = 16{,}8,\ n = 12,\ \gamma = 0{,}95.$

Respuesta $15{,}85 < a < 17{,}75.$

12. $s = 2{,}4,\ \overline{x}_m = 14{,}2,\ n = 9,\ \gamma = 0{,}99.$

Respuesta $11{,}512 < a < 16{,}888.$

13. Por los datos de 16 mediciones de igual precisión independientes de una magnitud física están hallados $\overline{x}_m = 23{,}161$ y $s = 0{,}400$. Se necesita estimar el valor real a de la magnitud que se mide y la precisión de las mediciones σ con la fiabilidad 0,95.

Respuesta $22{,}948 < a < 23{,}374;\ 0{,}224 < \sigma < 0{,}576.$

Capítulo diez y siete

METODOS DE CALCULO DE LAS CARACTERISTICAS GENERALES DE UNA MUESTRA

§ 1. Variantes condicionales

Supongamos que las variantes de la muestra están dispuestas en orden creciente, es decir, en forma de una serie de variación.

Las variantes que forman una progresión aritmética de diferencia h se llaman *equidistantes*.

Las variantes determinadas por la igualdad

$$u_i = \frac{x_i - C}{h},$$

donde C es un cero accidental (nuevo origen de lectura);
h es el paso, es decir, la diferencia entre dos variantes originales contiguas cualesquiera (nueva unidad de escala),

se llaman *condicionales*.

Los métodos simplificados de cálculo de las características generales de una muestra se basan en la sustitución de las variantes originales por las condicionales.

Demostraremos que si la serie de variación se compone de variantes equidistantes de paso h, las variantes condicionales son números enteros. En efecto, tomemos como cero accidental una variante arbitraria, por ejemplo x_m. En tal

caso

$$u_i = \frac{x_i - x_m}{h} = \frac{x_1 + (i-1)h - [x_1 + (m-1)h]}{h} = i - m.$$

En virtud de que i y m son números enteros, su diferencia $i - m = u_i$ también es un número entero.

Nota 1. Como cero accidental se puede tomar cualquier variante. La simpleza máxima de los cálculos se logra cuando se toma como cero accidental la variante que se encuentra aproximadamente en el centro de la serie de variación (de ordinario, tal variante tiene frecuencia máxima).

Nota 2. A la variante que se toma como cero accidental le corresponde la variante condicional, que es igual a cero.

Ejemplo. Hallar las variantes condicionales de una distribución estadística

variantes	23,6	28,6	33,6	38,6	43,6
frecuencias	5	20	50	15	10.

SOLUCION. Tomemos como cero accidental la variante 33,6 (esta variante se encuentra en el centro de la serie de variación).

Hallamos el paso

$$h = 28,6 - 23,6 = 5.$$

Hallamos la variante condicional

$$u_1 = \frac{x_1 - C}{h} = \frac{23,6 - 33,6}{5} = -2.$$

Análogamente obtenemos

$$u_2 = -1, \quad u_3 = 0, \quad u_4 = 1, \quad u_5 = 2.$$

Como podemos apreciar las variantes condicionales son pequeños números enteros. Indudablemente, operar con ellas es más sencillo que con las variantes originales.

§ 2. Momentos empíricos ordinarios, iniciales y centrales

Para calcular las características generales de una muestra conviene utilizar los momentos empíricos, determinados de manera análoga a los correspondientes momentos teóricos (cap. VIII, § 10). A diferencia de los momentos teóricos, los empíricos se calculan por los datos de observaciones.

Se llama *momento empírico ordinario de orden k* el valor medio de las diferencias $x_i - c$ de potencias k:

$$M'_k = \frac{\sum n_i (x_i - c)^k}{n},$$

donde x_i es la variante que se observa,
n_i es la frecuencia de la variante,
$n = \sum n_i$ es el volumen de la muestra,
c es un número constante arbitrario (cero accidental).

Se llama *momento empírico inicial de orden k* el momento ordinario de orden k cuando $c = 0$

$$M_k = \frac{\sum n_i x_i^k}{n}.$$

En particular,

$$M_1 = \frac{\sum n_i x_i}{n} = \overline{x}_m,$$

es decir, el momento empírico de primer orden es igual a la media muestral.

Se llama *momento empírico central de orden k* el momento ordinario de orden k para $c = \overline{x}_m$

$$m_k = \frac{\sum n_i (x_i - \overline{x}_m)^k}{n}.$$

En particular,

$$m_2 = \frac{\sum n_i (x_i - \overline{x}_m)^2}{n} = D_m, \qquad (*)$$

es decir, el momento empírico central de segundo orden es igual a la dispersión muestral.

Los momentos centrales se expresan fácilmente por los ordinarios (recomendamos a los lectores hacerlo individualmente):

$$m_2 = M'_2 - (M'_1)^2; \qquad (**)$$

$$\left. \begin{array}{r} m_3 = M'_3 - 3M'_2 M'_1 + 2(M'_1)^3; \\ m_4 = M'_4 - 4M'_3 M'_1 + 6M'_2 (M'_1)^2 - 3(M'_1)^4. \end{array} \right\} \qquad (***)$$

§ 3. Momentos empíricos condicionales. Obtención de los momentos centrales por los condicionales

El cálculo de los momentos centrales es bastante voluminoso. Para simplificar los cálculos se reemplazan las variantes originales por las condicionales.

Se llama *momento empírico condicional de orden k* el momento inicial de orden k, calculado para las variantes condicionales:

$$M_k^* = \frac{\sum n_i u_i^k}{n} = \frac{\sum n_i \left(\frac{x_i - c}{h}\right)^k}{n}.$$

En particular,

$$M_1^* = \frac{\sum n_i \left(\frac{x_i - c}{h}\right)}{n} = \frac{1}{h}\left[\frac{\sum n_i x_i}{n} - c\,\frac{\sum n_i}{n}\right] = \frac{1}{h}\left(\bar{x}_m - c\right).$$

De aquí

$$\bar{x}_m = M_1^* h + c. \qquad (*)$$

Por consiguiente, para hallar la media muestral es suficiente calcular el momento condicional de primer orden, multiplicarlo por h y sumar al resultado el cero accidental c.

Expresemos los momentos ordinarios por los condicionales:

$$M_k^* = \frac{1}{h^k}\cdot\frac{\sum n_i (x_i - c)^k}{n} = \frac{M_k'}{h^k}.$$

De aquí

$$M_k' = M_k^* h^k.$$

De este modo, para hallar el momento ordinario de orden k es suficiente multiplicar el momento condicional de igual orden por h^k.

Obtenidos los momentos ordinarios se hallan fácilmente los momentos centrales por las igualdades (**) y (***) del párrafo anterior. En suma, obtenemos fórmulas convenientes para los cálculos, que expresan los momentos centrales por los condicionales:

$$m_2 = [M_2^* - (M_1^*)^2]\,h^2; \qquad (**)$$

$$\left.\begin{aligned} m_3 &= [M_3^* - 3M_2^* M_1^* + 2(M_1^*)^3]\,h^3;\\ m_4 &= [M_4^* - 4M_3^* M_1^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4]\,h^4. \end{aligned}\right\} \qquad (***)$$

En particular, en virtud de (**) y la correlación (*) del párrafo precedente obtenemos la fórmula para calcular la dispersión muestral por los momentos condicionales de primer y segundo órdenes

$$D_m = [M_2^* - (M_1^*)^2]\, h^2. \qquad (****)$$

La técnica de los cálculos de los momentos centrales por las condicionales se describe más adelante.

§ 4. Método de los productos del cálculo de la media y la dispersión muestrales

El método de los productos da un medio conveniente de cálculo de los momentos condicionales de diferentes órdenes de una serie de variación con variantes equidistantes. Sabiendo los momentos condicionales se hallan fácilmente los momentos empíricos iniciales y centrales que nos interesan. En particular, por el método de los productos es conveniente calcular la media muestral y la dispersión muestral. Es útil emplear la tabla de cálculo formada del modo siguiente:

1) en la primera columna de la tabla se escriben las variantes muestrales (originales), disponiéndolas en orden creciente;

2) en la segunda columna se escriben las frecuencias de la variante; se ponen todas las frecuencias y su suma (volumen de la muestra n) se coloca en la casilla inferior de la columna;

3) en la tercera columna se escriben las variantes condicionales $u_i = \frac{x_i - C}{h}$; además, como cero accidental C se toma la variante de frecuencia máxima y se supone h igual a la diferencia entre dos variantes contiguas cualesquiera; prácticamente la tercera columna se llena así: en la casilla de la línea que contiene la frecuencia máxima se escribe el 0; en las casillas encima del cero se escriben sucesivamente -1, -2, -3, etc., y debajo del cero 1, 2, 3, etc;

4) se multiplican las frecuencias por las variantes condicionales y sus productos $n_i u_i$ se escriben en la cuarta columna; sumando todos los números obtenidos, su suma $\sum n_i u_i$ se pone en la casilla inferior de la columna;

5) se multiplican las frecuencias por los cuadrados de las variantes condicionales y sus productos $n_i u_i^2$ se escriben en la quinta columna; sumando todos los números obtenidos, su suma $\sum n_i u_i^2$ se coloca en la casilla inferior de la columna;

6) se multiplican las frecuencias por los cuadrados de las variantes condicionales aumentados en una unidad, y los productos $n_i (u_i + 1)^2$ se escribe en la sexta columna; sumando los números obtenidos, su suma $\sum n_i (u_i + 1)^2$ se coloca en la casilla inferior de la columna.

Nota 1. Conviene sumar por separado los números negativos de la cuarta columna (su suma A_1 se escribe en la casilla de la línea que contiene la frecuencia máxima) y los positivos (su suma A_2 se escribe en la penúltima casilla de la columna); en tal caso $\sum n_i u_i = A_1 + A_2$.

Nota 2. Al calcular los productos $n_i u_i^2$ de la quinta columna, los números $n_i u_i$ de la cuarta columna conviene multiplicarlos por u_i.

Nota 3. La sexta columna sirve para verificar los cálculos: si la suma $\sum n_i (u_i + 1)^2$ resulta igual a la suma $\sum n_i u_i^2 + 2 \sum n_i u_i + n$ (como debe ser de acuerdo con la igualdad $\sum n_i (u_i + 1)^2 = \sum n_i u_i^2 + 2 \sum n_i u_i + n$), los cálculos son correctos.

Nota 4. Como cero accidental se puede tomar cualquier variante, es decir, no es obligatorio tomar la variante de frecuencia máxima, como se dijo en el p. 3. Por ejemplo, si la variante que tiene la frecuencia máxima se encuentra en las primeras o últimas líneas de la «columna x_i», lo *más conveniente es tomar como cero accidental la variante que se encuentra aproximadamente en el centro de la columna.*

Después de completar la tabla de cálculo y verificar la corrección de los cálculos, se procede al cómputo de los momentos condicionales:

$$M_1^* = \frac{\sum n_i u_i}{n}, \qquad M_2^* = \frac{\sum n_i u_i^2}{n}.$$

Finalmente se calculan la media y la dispersión muestrales por las fórmulas (*) y (****) § 3:

$$\overline{x}_m = M_1^* \cdot h + C,$$
$$D_m = [M_2^* - (M_1^*)^2] h^2.$$

Ejemplo. Hallar por el método de los productos la media y la dispersión muestrales de la siguiente distribución estadística:

variantes:	10,2	10,4	10,6	10,8	11,0	11,2	11,4	11,6	11,8	12,0
frecuencia:	2	3	8	13	25	20	12	10	6	1

SOLUCION. Formamos la tabla de cálculo, para ello:

1) escribimos las variantes en la primera columna;

2) escribimos las frecuencias en la segunda columna; la suma de las frecuencias (100) la colocamos en la casilla inferior de la columna;

3) como cero accidental tomamos la variante 11,0 (esta variante tiene la frecuencia máxima); en la casilla de la tercera columna, correspondiente a la línea que contiene la frecuencia máxima, escribimos el 0; sobre el cero sucesivamente −1, −2, −3, −4, y debajo del cero 1, 2, 3, 4, 5;

4) los productos de las frecuencias por las variantes condicionales los escribimos en la cuarta columna; hallamos por separado la suma (−46) de números negativos y por separado la suma de números positivos (103); sumando estos números, su total (57) lo colocamos en la casilla inferior de la columna;

5) los productos de las frecuencias por los cuadrados de las variantes condicionales los escribimos en la quinta columna; la suma de los números de la columna (383) la colocamos en la casilla inferior de la columna;

6) los productos de las frecuencias por los cuadrados de las varientes condicionales incrementados en la unidad, los escribimos en la sexta columna de control; la suma (597) de los números de la columna la colocamos en la casilla inferior de la columna.

Como resultado obtenemos la tabla de cálculo 7.

Verificación: $\sum n_i u_i^2 + 2\sum n_i u_i + n = 383 + 2\cdot 57 + 100 = 597$.

$$\sum n_i (u_i + 1)^2 = 597.$$

Los cálculos son correctos.

Calculemos los momentos condicionales de primer y segundo órdenes

$$M_1^* = \frac{\sum n_i u_i}{n} = \frac{57}{100} = 0{,}57;$$

$$M_2^* = \frac{\sum n_i u_i^2}{n} = \frac{383}{100} = 3{,}83.$$

Hallamos el paso: $h = 10{,}4 - 10{,}2 = 0{,}2$.

Calculamos la media y la dispersión muestrales buscadas:

$$\overline{x}_m = M_1^* \cdot h + C = 0{,}57 \cdot 0{,}2 + 11{,}0 = 11{,}1;$$

$$D_m = [M_2^* - (M_1^*)^2] \cdot h^2 = [3{,}83 - (0{,}57)^2] \cdot 0{,}2^2 = 0{,}14.$$

Tabla 7

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i+1)^2$
10,2	2	−4	−8	32	18
10,4	3	−3	−9	27	12
10,6	8	−2	−16	32	8
10,8	13	−1	−13	13	0
11,0	25	0	$A_1 = -46$		25
11,2	20	1	20	20	80
11,4	12	2	24	48	108
11,6	10	3	30	90	160
11,8	6	4	24	96	150
12,0	1	5	5	25	36
			$A_2 = 103$		
	$n=100$		$\sum n_i u_i = 57$	$\sum n_i u_i^2 = 383$	$\sum n_i (u_i+1)^2 = 597$

§ 5. Reducción de las variantes originales a equidistantes

Antes se expuso el método de cálculo de las características muestrales para las variantes equidistantes. Como regla, en la práctica los datos de observaciones no serán números equidistantes. Naturalmente, surge la pregunta: ¿acaso no se podría reducir el cálculo de los valores obser-

vados del carácter al caso de las variantes equidistantes mediante la correspondiente transformación? Resulta que se puede. Con este propósito el intervalo, en el que se encuentran todos los valores observados del carácter (variantes originales), se divide en varios intervalos parciales iguales. (Prácticamente en cada intervalo parcial deben caer no menos de 8—10 variantes originales.) A continuación se hallan los centros de los intervalos parciales, los que forman precisamente la sucesión de variantes equidistantes.

Como frecuencia de cada «nueva» variante (centro del intervalo parcial) se utiliza el número total de variantes originales que caen en el intervalo parcial correspondiente.

Está claro que el reemplazo de las variantes originales por los centros de los intervalos parciales da lugar a errores (las variantes originales de la mitad izquierda del intervalo parcial serán incrementadas, en tanto que las variantes de la mitad derecha serán disminuidas), sin embargo estos errores serán anulados, puesto que tienen distintos signos.

Ejemplo. Un conjunto muestral de volumen $n = 100$ está prefijado por la tabla 8.

Tabla 8

x_i	n_i	x_i	n_i	x_i	n_i
1,00	1	1,19	2	1,37	6
1,03	3	1,20	4	1,38	2
1,05	6	1,23	4	1,39	1
1,06	4	1,25	8	1,40	2
1,08	2	1,26	4	1,44	3
1,10	4	1,29	4	1,45	3
1,12	3	1,30	6	1,46	2
1,15	6	1,32	4	1,49	4
1,16	5	1,33	5	1,50	2

Formar la distribución de las variantes equidistantes.

SOLUCION. Dividimos el intervalo 1,00—1,50, por ejemplo, en los 5 intervalos parciales siguientes: 1,00 — 1,10; 1,10—1,20; 1,20—1,30; 1,30—1,40; 1,40—1,50.

Tomando los centros de los intervalos parciales como nuevas

variantes y_i, obtenemos las variantes equidistantes:

$y_1 = 1{,}05$; $y_2 = 1{,}15$; $y_3 = 1{,}25$; $y_4 = 1{,}35$; $y_5 = 1{,}45$.

Hallamos la frecuencia de la variante y_1:

$$n_1 = 1 + 3 + 6 + 4 + 2 + \frac{4}{2} = 18.$$

(Puesto que la variante original 1,10 es al mismo tiempo final del primer intervalo parcial y principio del segundo, la frecuencia 4 de esta variante está dividida en partes iguales entre ambos intervalos parciales.)

Hallamos la frecuencia de la variante y_2:

$$n_2 = \frac{4}{2} + 3 + 6 + 5 + 2 + \frac{4}{2} = 20.$$

Análogamente calculamos las frecuencias de las demás variantes:

$$n_3 = 25; \quad n_4 = 22; \quad n_5 = 15.$$

En conclusión obtenemos la siguiente distribución de las variantes equidistantes:

y_i	1,05	1,15	1,25	1,35	1,45
n_i	18	20	25	22	15.

A título de ejercicio recomendamos al lector comprobar que las medias y las dispersiones muestrales calculadas por las variantes originales y equidistantes resultan, respectivamente, iguales a:

$$\bar{x}_m = 1{,}250; \qquad \bar{y}_m = 1{,}246;$$
$$D_x = 0{,}018; \qquad D_y = 0{,}017.$$

Como podemos apreciar, el reemplazo de las variantes originales por las equidistantes no dio lugar a errores sustanciales; en este caso se reduce bastante el volumen del cálculo.

§ 6. Frecuencias empíricas y de igualación (teóricas)

A. Distribución discreta

Examinemos una magnitud aleatoria discreta X cuya ley de distribución está desconocida. Supongamos que se han realizado n experimentos, en los cuales la magnitud X tomó n_1 veces el valor x_1, n_2 veces el valor x_2, . . ., n_k veces el valor x_k; asimismo $\sum n_i = n$.

Se llaman *frecuencias empíricas* las frecuencias observadas prácticamente n_i.

Admitamos tener fundamentos para suponer que la magnitud, que se estudia X, está distribuida por cierta ley determinada. Para verificar si esta suposición concuerda con los datos de las observaciones *se calculan* las frecuencias de los valores a observar, es decir, *se halla teóricamente* cuántas veces la magnitud X tendría que tomar cada uno de los valores que se observan si ella está distribuida por una ley hipotética.

A diferencia de las frecuencias empíricas observadas prácticamente, las frecuencias n'_i halladas teóricamente (por cálculo) se llaman *de igualación* (*teóricas*).

Las frecuencias de igualación se hallan por la igualdad

$$n'_i = nP_i,$$

donde n es el número de pruebas,

P_i es la probabilidad del valor que se observa x_i, calculada suponiendo que X tiene una distribución hipotética. Esta fórmula se deduce del teorema de la esperanza matemática del número de apariciones de un suceso en pruebas independientes (cap. VII, § 5).

De este modo, *la frecuencia de igualación del valor observado x_i de distribución discreta es igual al producto del número de pruebas por la probabilidad de este valor observado.*

Ejemplo. Como resultado del experimento consistente de $n = 520$ pruebas, en cada una de las cuales se registró el número x_i de apariciones de cierto suceso, se ha obtenido la siguiente distribución empírica

val. obser.	x_i	0	1	2	3	4	5	6	7
frec. emp.	n_i	120	167	130	69	27	5	1	1.

Hallar las frecuencias de igualación n'_i, suponiendo que la magnitud aleatoria X (conjunto general) está distribuida por la ley de Poisson.

SOLUCIÓN. Sabemos que el parámetro λ, por el que se determina la distribución de Poisson, es igual a la esperanza matemática de esta distribución. Puesto que como estimación de la esperanza matemática se toma la media muestral (cap. XVI, § 5), como estimación de λ se puede tomar la media muestral $\overline{x}_m$. Por los datos se halla fácilmente que la media muestral es igual a 1,5; por lo tanto, se puede tomar $\lambda = 1,5$.

De esta manera, la fórmula de Poisson

$$P_n(k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}.$$

toma la forma:

$$P_{520}(k) = \frac{1,5^k \cdot e^{-1,5}}{k!}.$$

Utilizando esta fórmula hallamos las probabilidades $P_{520}(k)$ para $k = 0, 1, 2, 3, 4, 5, 6, 7$ (para simpleza de la notación en adelante omitimos el subíndice 520): $P(0) = 0{,}22313$, $P(1) = 0{,}33469$, $P(2) = 0{,}251021$, $P(3) = 0{,}125511$, $P(4) = 0{,}047066$, $P(5) = 0{,}014120$, $P(6) = 0{,}003530$, $P(7) = 0{,}000755$.

Hallamos las frecuencias de igualación (los resultados están redondeados hasta la unidad):

$$n'_1 = n \cdot P(0) = 520 \cdot 0{,}22313 = 116,$$

$$n'_2 = n \cdot P(1) = 520 \cdot 0{,}33469 = 174.$$

Análogamente se hallan las demás frecuencias de igualación. En resumen obtenemos:

frecuencia emp.	123	167	130	69	27	5	1	1
frecuencia de igua.	116	174	131	65	25	7	2	0.

La divergencia relativamente pequeña de las frecuencias empíricas y de igualación confirma la suposición de que la distribución examinada obedece a la ley de Poisson.

Cabe hacer notar que si la dispersión muestral se calcula por los datos de la distribución, resulta que ella es igual a la media muestral, es decir, igual a 1,5. Esto corrobora una vez más la suposición hecha, ya que para la distribución de Poisson $\lambda = M(X) = D(X)$.

B. Distribución continua

En el caso de la distribución continua, las probabilidades de los valores posibles individuales son iguales a cero (cap. X, § 2, corolario 2). Por eso todo el intervalo de valores posibles se divide en k intervalos no intersecables y se calculan las probabilidades P_i de que X caiga en el i-ésimo intervalo parcial; a continuación, al igual que para la distribución discreta, se multiplica el número de pruebas por estas probabilidades.

Así pues, *las frecuencias de igualación de una distribución continua* se hallan por la igualdad

$$n'_i = nP_i,$$

donde n es el número de pruebas,

P_i es probabilidad de que X caiga en el i-ésimo intervalo parcial, calculada admitiendo que X tiene la distribución supuesta.

En particular, si existen fundamentos para suponer que la magnitud aleatoria X (conjunto general) está distribuida *normalmente*, las frecuencias de igualación pueden ser halladas por la fórmula

$$n'_i = \frac{nh}{\sigma_m}\varphi(u_i), \tag{*}$$

donde n es el número de pruebas (volumen de la muestra),

h es la longitud del intervalo parcial,

σ_m es la desviación cuadrática media muestral,

$u_i = \frac{x_i - \bar{x}_m}{\sigma_m}$ (x_i es el centro del i-ésimo intervalo parcial),

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

En el § 7 se da un ejemplo de aplicación de la fórmula (*).

Explicación. Aclaremos el origen de la fórmula (*). Escribimos la función diferencial de la distribución normal general

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \tag{**}$$

Para $a = 0$ y $\sigma = 1$ obtenemos la función diferencial de la distribución normada

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

o bien, cambiando la designación del argumento,

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

Poniendo $u = \frac{x-a}{\sigma}$, tendremos

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (***)$$

Comparando (**) y (***), deducimos que

$$f(x) = \frac{1}{\sigma} \varphi(u).$$

Si la esperanza matemática a y la desviación cuadrática media σ son desconocidas, como estimaciones de estos parámetros se toman respectivamente la media muestral $\bar{x}_m$ y la desviación cuadrática media muestral σ_m (cap. XVI, §§ 5, 9). En tal caso,

$$f(x) = \frac{1}{\sigma_m} \varphi(u),$$

donde $u = \frac{x - \bar{x}_m}{\sigma_m}$.

Supongamos que x_i es el centro del i-ésimo intervalo (en los que está dividido el conjunto de todos los valores observados de la magnitud aleatoria X normalmente distribuida) de longitud h. Entonces la probabilidad de que X caiga en ese intervalo es aproximadamente igual al producto de la longitud del intervalo por el valor de la función diferencial $f(x)$ en cualquier punto del intervalo y, en particular, para $x = x_i$ (cap. XI, § 5):

$$P_i = hf(x_i) = h \cdot \frac{1}{\sigma_m} \varphi(u_i).$$

Por lo tanto, la frecuencia de igualación es

$$n'_i = nP_i = \frac{nh}{\sigma_m} \varphi(u_i).$$

donde

$$u_i = \frac{x_i - \bar{x}_m}{\sigma_m}.$$

Hemos obtenido la fórmula (*).

§ 7. Trazado de la curva de Gauss por datos experimentales

Uno de los métodos de construcción de una curva normal o de Gauss por los datos de observaciones consiste en lo siguiente:

1) hallar $\bar{x}_m$ y σ_m, por ejemplo, por el método de los productos;

2) hallar las ordenadas y_i (frecuencias de igualación) de la curva teórica por la fórmula $y_i = \frac{n \cdot h}{\sigma_m} \cdot \varphi(u_i)$, donde n es la suma de las frecuencias observada, h es la diferencia entre dos variantes contiguas $u_i = \frac{x_i - \bar{x}_m}{\sigma_m}$ y $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$;

3) trazar los puntos (x_i, y_i) en un sistema de coordenadas cartesianas y unirlos por una curva suave.

La proximidad de las frecuencias de igualación a las observadas confirma que la admisión de que el carácter estudiado está distribuido normalmente, es correcta.

Ejemplo. Trazar la curva de Gauss por la distribución dada:

variante	x_i	15	20	25	30	35	40	45	50	55
frecuencia	n_i	6	13	38	74	106	85	30	10	4.

SOLUCION. Utilizando el método de los productos (§ 4) hallamos $\bar{x}_m = 34{,}7$, $\sigma_m = 7{,}38$.

Calculamos las frecencias de igualación (véase la tabla 9).

Tabla 9

x_i	n_i	$x_i - \bar{x}_m$	$u_i = \frac{x_i - \bar{x}_m}{\sigma_m}$	$\varphi(u_i)$	$y_i = \frac{n \cdot h}{\sigma_m} \cdot \varphi(u_i) = 248 \cdot \varphi(u_i)$
15	6	−19,7	−2,67	0,0113	3
20	13	−14,7	−1,99	0,0551	14
25	38	−9,7	−1,31	0,1691	42
30	74	−4,7	−0,63	0,3271	82
35	106	0,3	0,05	0,3984	99
40	85	5,3	0,73	0,3056	76
45	30	10,3	1,41	0,1476	37
50	10	15,3	2,09	0,0449	11
55	4	20,3	2,77	0,0086	2
	$n = 366$				$\sum y_i = 366$

En la fig. 22 se han trazado la curva de Gauss (teórica) por las frecuencias de igualación (señaladas con círculos) y el polígono de las frecuencias observadas (señaladas con

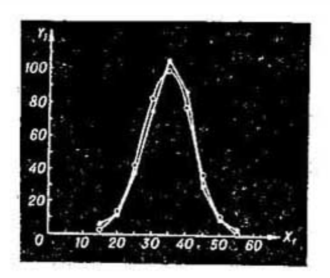


Fig. 22.

cruces). Comparando las gráficas se aprecia que la curva teórica construida refleja satisfactoriamente los datos de las observaciones.

Para considerar con más certeza que los datos de las observaciones evidencian la distribución normal del carácter, se utilizan reglas especiales (denominadas criterios de concordancia), cuyas nociones el lector las encontrará más adelante (cap. XIX, § 22).

§ 8. Estimación de la desviación de una distribución empírica respecto de la normal. Asimetría y exceso

Para estimar la desviación de una distribución empírica respecto de la normal (distribución de Gauss) se utilizan distintas características, entre las cuales se encuentran la asimetría y el exceso. Las definiciones de estas características son análogas a las definiciones de asimetría y exceso de la distribución teórica (cap. XII, § 9).

La asimetría de una distribución empírica se determina por la igualdad:

$$a_s = \frac{m_3}{\sigma_m^3},$$

donde m_3 es el momento empírico central de tercer orden (§ 2).

El exceso de una distribución empírica se determina por la igualdad

$$e_k = \frac{m_4}{\sigma_m^4} - 3,$$

donde m_4 es el momento empírico central de cuarto orden.

Los momentos m_3 y m_4 conviene calcularlos por el método de los productos (§ 4), utilizando la fórmula (***) del § 3.

Ejemplo. Hallar la asimetría y el exceso de la distribución empírica:

variante	10,2	10,4	10,6	10,8	11,0	11,2	11,4	11,6	11,8	12,0
frecuencia	2	3	8	13	25	20	12	10	6	1

SOLUCIÓN. Utilizamos el método de los productos, para lo cual componemos la tabla de cálculo. Puesto que en el § 4 ya se indicó como se llenan las columnas 1—5 de la tabla, nos limitaremos a breves aclaraciones: para llenar la columna 6 conviene multiplicar los números de cada línea de las columnas 3 y 5; para llenar la columna 7 conviene multiplicar los números de cada línea de las columnas 3 y 6. La columna 8 sirve para controlar los cálculos por la igualdad:

$$\sum n_i (u_i+1)^4 = \sum n_i u_i^4 + 4\sum n_i u_i^3 + 6\sum n_i u_i^2 + 4\sum n_i u_i + n.$$

Damos la tabla de cálculo 10.

Verificación: $\sum n_i (u_i+1)^4 = 9141$;

$$\sum n_i u_i^4 + 4\sum n_i u_i^3 + 6\sum n_i u_i^2 + 4\sum n_i u_i + n =$$
$$= 4079 + 4\cdot 609 + 6\cdot 383 + 4\cdot 57 + 100 = 9141.$$

La coincidencia de las sumas nos demuestra que los cálculos son correctos.

En el ejemplo del § 4 se halló para la distribución examinada:

$$M_1^* = 0,57; \quad M_2^* = 3,83; \quad D_m = 0,14,$$

por lo tanto,

$$\sigma_m = \sqrt{0,14}.$$

Hallamos los momentos condicionales de tercer y cuarto órdenes:

$$M_3^* = \frac{\sum n_i u_i^3}{n} = \frac{609}{100} = 6{,}09; \quad M_4^* = \frac{\sum n_i u_i^4}{n} = \frac{4079}{100} = 40{,}79.$$

Tabla 10

1	2	3	4	5	6	7	8
x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	$n_i (u_i + 1)^4$
10,2	2	—4	—8	32	—128	512	162
10,4	3	—3	—9	27	—81	243	48
10,6	8	—2	—16	32	—64	128	8
10,8	13	—1	—13	13	—13	13	—
11,0	25	0	—46		—286		25
11,2	20	1	20	20	20	20	320
11,4	12	2	24	48	96	192	972
11,6	10	3	30	90	270	810	2560
11,8	6	4	24	96	384	1536	3750
12,0	1	5	5	25	125	625	1296
			103		895		
	$n = 100$		$\sum n_i u_i = 57$	$\sum n_i u_i^2 = 383$	$\sum n_i u_i^3 = 609$	$\sum n_i u_i^4 = 4079$	$\sum n_i (u_i + 1)^4 = 9141$

Hallamos los momentos empíricos centrales de tercer y cuarto órdenes:

$$m_3 = [M_3^* - 3M_1^* M_2^* + 2\,(M_1^*)^3]\,h^3 =$$
$$= [6{,}09 - 3 \cdot 0{,}57 \cdot 3{,}83 + 2 \cdot (0{,}57)^3] \cdot 0{,}2^3 = -0{,}0007;$$
$$m_4 = [M_4^* - 4M_1^* M_3^* + 6\,(M_1^*)^2\,M_2^* - 3\,(M_1^*)^4]\,h^4 =$$
$$= [40{,}79 - 4 \cdot 0{,}57 \cdot 6{,}09 + 6\,(0{,}57)^2 \cdot 3{,}83 - 3 \cdot (0{,}57)^4] \cdot 0{,}2^4 = 0{,}054.$$

Hallamos la asimetría y el exceso:

$$a_s = \frac{m_3}{\sigma_{in}^3} = -\frac{0{,}0007}{(\sqrt{0{,}14})^3} = -0{,}01;$$
$$e_k = \frac{m_4}{\sigma_{in}^4} - 3 = \frac{0{,}054}{(\sqrt{0{,}14})^4} - 3 = -0{,}24.$$

Nota. En el caso de pequeñas muestras hay que tratar con cuidado las estimaciones de la asimetría y el exceso y hallar la exactitud de estas estimaciones (véase N. V. Smirnov e I. V. Dunin-Barkovski, Curso de teoría de las probabilidades y estadística matemática. Ed. Nauka 1965, pág. 277).

Problemas

En los problemas 1—2 se dan las variantes muestrales y sus frecuencias. Hallar la media y la dispersión muestrales utilizando el método de los productos.

1.

x_i	10,3	10,5	10,7	10,9	11,1	11,3	11,5	11,7	11,9	12,1
n_i	4	7	8	10	25	15	12	10	4	5

Respuesta $\bar{x}_m = 11,19$, $D_m = 0,19$.

2.

x_i	83	85	87	89	91	93	95	97	99	101
n_i	6	7	12	15	30	10	8	6	4	2.

Respuesta $\bar{x}_m = 90,72$, $D_m = 17,20$.

3. Hallar la asimetría y el exceso de la distribución empírica

x_i	10,6	10,8	11,0	11,2	11,4	11,6	11,8
n_i	5	10	17	30	20	12	6.

Respuesta $a_s = -0,0006$, $e_k = 0,00004$.

Capítulo diez y ocho

ELEMENTOS DE LA TEORIA DE LA CORRELACION

§ 1. Dependencias funcional, estadística y de correlación

En muchos problemas hay que establecer y estimar la dependencia de una magnitud aleatoria Y a estudiar respecto de una o varias otras magnitudes. Examinemos al principio la dependencia de Y respecto de una magnitud aleatoria (o prevista) X y luego respecto de varias magnitudes (§ 15).

Dos magnitudes aleatorias pueden estar relacionadas por una dependencia funcional (cap. XII, § 10), o bien por una dependencia de otro género, llamada estadística, o pueden ser independientes.

Raramente se realiza una dependencia funcional rigurosa, ya que ambas magnitudes o una de ellas están expuestas a la acción de factores aleatorios; además, entre ellos pueden haber comunes para ambas magnitudes (por «comunes» aquí se sobreentienden los factores que actúan tanto sobre Y, como sobre X). En este caso surge una dependencia estadística.

Por ejemplo, si Y depende de los factores aleatorios

$$Z_1, Z_2, V_1, V_2,$$

y X depende de los factores aleatorios

$$Z_1, Z_2, U_1,$$

entre Y y X existe una dependencia estadística, puesto que entre los factores aleatorios hay comunes, es decir, Z_1 y Z_2.

La dependencia se llama *estadística* cuando la variación de una de las magnitudes da lugar a la alteración de la distribución de la otra. En particular, la dependencia estadística se manifiesta en que, al variar una de las magnitudes, se altera el valor medio de la otra; en este caso la dependencia estadística se llama *de correlación*.

Damos un ejemplo de una magnitud aleatoria Y que no está vinculada funcionalmente con la magnitud X, sino lo está correlativamente. Supongamos que Y es una cosecha de granos, X es la cantidad de abonos. De iguales parcelas de tierra con iguales cantidades de abonamiento se obtiene diferente cosecha, es decir, Y no es función de X. Esto se debe a la acción de factores aleatorios (precipitaciones, temperatura ambiente, etc.). Al mismo tiempo, como lo muestra la experiencia, la cosecha *media* es función de la cantidad de abonos, es decir, Y está vinculada con X por una dependencia de correlación.

§ 2. Medias condicionales. Dependencia de correlación

Precisemos la definición de la dependencia de correlación, para lo cual introducimos el concepto de media condicional.

Supongamos que se estudia el enlace entre la magnitud aleatoria Y y la magnitud aleatoria X. Admitamos que a cada valor de X corresponden varios valores de Y. Por ejemplo, supongamos que para $x_1 = 2$ la magnitud Y toma los valores: $y_1 = 5$, $y_2 = 6$, $y_3 = 10$. Hallemos la media aritmética de estos números:

$$\bar{y}_2 = \frac{5+6+10}{3} = 7.$$

El número $\bar{y}_2$ se llama media condicional; el pequeño trazo sobre la letra sirve para designar la media aritmética y el número 2 indica que se consideran los valores de Y que corresponden a $x_1 = 2$.

Conforme al ejemplo del párrafo anterior estos datos se pueden interpretar así: en cada una de las tres parcelas idénticas se han esparcido 2 unidades de abono y se han obtenido respectivamente 5, 6 y 10 unidades de grano; la cosecha media es de 7 unidades correspondientes.

Se llama *media condicional* $\bar{y}_x$ la media aritmética de los valores de Y correspondientes al valor de $X = x$.

Si a cada valor de x corresponde un valor de la media condicional, evidentemente, la media condicional es una función de x; en este caso se dice que la magnitud aleatoria Y depende de X correlativamente.

Se llama *dependencia de correlación de* Y respecto de X la dependencia funcional de la media condicional $\bar{y}_x$ respecto de x:

$$\bar{y}_x = f(x). \qquad (*)$$

La ecuación (*) se llama *ecuación de regresión* de Y en X; la función $f(x)$ se llama regresión de Y en X, y su gráfica, *línea de regresión* de Y en X.

Análogamente se determina la media condicional $\bar{x}_y$ y la dependencia de correlación de X respecto de Y.

Se llama *media condicional* $\bar{x}_y$ la media aritmética de los valores de X correspondientes a $Y = y$.

Se llama *dependencia de correlación* de X respecto de Y la dependencia funcional de la media condicional $\bar{x}_y$ respecto de y:

$$\bar{x}_y = \varphi(y). \qquad (**)$$

La ecuación (**) se llama *ecuación de regresión* de X en Y; la función $\varphi(y)$ se llama *regresión* de X en Y y su gráfica, *línea de regresión* de X en Y.

§ 3. Dos problemas fundamentales de la teoría de la correlación

El primer problema de la teoría de la correlación consiste en establecer la forma del enlace de correlación, es decir, el tipo de función de regresión (lineal, cuadrática, exponencial etc.). En la mayoría de los casos las funciones de regresión son lineales. Si ambas funciones de regresión $f(x)$ y $\varphi(y)$ son lineales, la correlación se llama *lineal*; en caso contrario, *no lineal*. Evidentemente, para la correlación lineal ambas líneas de regresión son rectas.

El segundo problema de la teoría de la correlación es estimar la estrechez (fuerza) del enlace de correlación. La fuerza de la dependencia de correlación de Y respecto de X se estima por la magnitud de la dispersión de los valores de Y alrededor de la media condicional $\bar{y}_x$. Una gran dispersión denota la débil dependencia de Y respecto de X o la falta de dependencia. Una pequeña dispersión indica la existencia de una dependencia suficientemente fuerte: incluso es posible que Y y X estén relacionados funcionalmente, pero por acción de factores aleatorios secundarios este enlace resulte interrumpido, debido a lo cual para igual valor de x la magnitud Y toma distintos valores.

Análogamente (por la magnitud de la dispersión de los valores de X alrededor de la media condicional $\bar{x}_y$) se estima la fuerza del enlace de correlación de X respecto de Y.

§ 4. Hallazgo de los parámetros de la ecuación muestral de la recta de regresión por datos no agrupados

Supongamos que los caracteres cuantitativos de X e Y están vinculados por una dependencia de correlación lineal. En este caso ambas líneas de regresión serán rectas.

Admitamos que para hallar las ecuaciones de estas rectas se han efectuado n experimentos independientes, en virtud de los cuales se han obtenido n pares de números:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

Puesto que los pares de números observados se puede considerar como una muestra aleatoria del conjunto general de

todos los valores posibles de la magnitud aleatoria (X, Y), las magnitudes y las ecuaciones halladas por estos datos se llaman *muestrales*.

Para certeza vamos a hallar la ecuación muestral de la recta de la regresión de Y en X.

Examinemos un caso elemental: distintos valores de x del carácter X y correspondientes a ellos los valores de y del carácter Y se han observado una sola vez cada uno. Es evidente que no hay necesidad de agrupar los datos. Tampoco hace falta utilizar el concepto de media condicional, por eso la ecuación buscada

$$y_x = kx + b$$

se puede escribir así:

$$Y = kx + b.$$

El coeficiente angular de la recta de la regresión de Y en X se llama *coeficiente de la regresión muestral* de Y en X y se designa por ρ_{yx}.

Así pues, vamos a hallar la ecuación muestral de la recta de la regresión de Y en X del tipo:

$$Y = \rho_{yx}x + b. \qquad (*)$$

Tratemos de escoger los parámetros ρ_{yx} y b de manera que los puntos (x_1, y_1) (x_2, y_2), . . ., (x_n, y_n) trazados en el plano XOY por los datos de las observaciones, se encuentren lo más cercano posible de la recta (*).

Precisemos el sentido de este requisito. Llamamos desviación la diferencia

$$Y_i - y_i \ (i = 1, 2, \ldots, n),$$

donde Y_i es una ordenada calculada por la ecuación (*), correspondiente al valor observado de x_i;

y_i es la ordenada observada, correspondiente a x_i.

Elegimos los parámetros ρ_{yx} y b de manera que la suma de los cuadrados de las desviaciones sea mínima (en esto está la esencia del método de los cuadrados mínimos).

Dado que cada desviación depende de los parámetros buscados, también la suma de los cuadrados de las desviaciones es una función F de estos parámetros (provisoriamente

en lugar de ρ_{yx} escribiremos ρ):

$$F(\rho, b) = \sum_{i=1}^{n} (Y_i - y_i)^2,$$

o bien

$$F(\rho, b) = \sum_{i=1}^{n} (\rho x_i + b - y_i)^2.$$

Para hallar el mínimo igualamos a cero las correspondientes derivadas parciales

$$\frac{\partial F}{\partial \rho} \quad 2 \sum_{i=1}^{n} (\rho x_i + b - y_i) x_i = 0;$$

$$\frac{\partial F}{\partial b} = 2 \sum_{i=1}^{n} (\rho x_i + b - y_i) = 0.$$

Mediante transformaciones elementales obtenemos un sistema de dos ecuaciones lineales con respecto a ρ y b.*

$$\left(\sum x^2\right) \rho + \left(\sum x\right) b = \sum xy; \quad \left(\sum x\right) \rho + nb = \sum y. \qquad (**)$$

Resolviendo este sistema hallamos los parámetros buscados:

$$\rho_y = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - \left(\sum x\right)^2}; \qquad (***)$$

$$b = \frac{\sum x^2 \cdot \sum y - \sum x \cdot \sum xy}{n \sum x^2 - \left(\sum x\right)^2}.$$

Análogamente podemos hallar la ecuación muestral de la recta de la regresión de X en Y:

$$\bar{x}_y = \rho_{xy} x + C,$$

donde ρ_{xy} es el coeficiente de la regresión muestral de X en Y.

Ejemplo. Hallar la ecuación muestral de la recta de la regresión de Y en X por los datos de $n = 5$ observaciones:

x	1,00	1,50	3,00	4,50	5,00;
y	1,25	1,40	1,50	1,75	2,25.

---

* Para simpleza, en lugar de $\sum_{i=1}^{n}$ escribiremos sólo $\sum$.

SOLUCIÓN. Componemos la tabla de cálculo 11.

Tabla 11

x_i	y_i	x_i^2	$x_i \cdot y_i$
1,00	1,25	1,00	1,250
1,50	1,40	2,25	2,100
3,00	1,50	9,00	4,500
4,50	1,75	20,25	4,875
5,00	2,25	25,00	11,250
$\sum x_i=15$	$\sum y_i=8,15$	$\sum x_i^2=57,50$	$\sum x_i y_i=26,975$

Hallamos los parámetros buscados, para lo cual sustituimos las sumas calculadas por la tabla en las correlaciones (***):

$$\rho_{yx}=\frac{5\cdot 26,975-15\cdot 8,15}{5\cdot 57,5-15^2}=0,202;$$

$$b=\frac{57,5\cdot 8,15-15\cdot 26,975}{62,5}=1,024.$$

Escribimos la ecuación buscada de la regresión:

$$Y=0,202x+1,024.$$

Para tener una idea hasta qué punto los valores de Y_i, calculados por esta ecuación, concuerdan bien con los valores observados de y_i, hallamos las desviaciones Y_i-y_i. Los resultados de los cálculos están llevados a la tabla 12.

Tabla 12

x_i	Y_i	y_i	$Y_i - y_i$
1,00	1,226	1,25	−0,024
1,50	1,327	1,40	−0,073
3,00	1,630	1,50	0,130
4,50	1,933	1,75	0,083
5,00	2,034	2,25	−0,216

Como se aprecia de la tabla, no todas las desviaciones son suficientemente pequeñas. Esto se debe al pequeño número de observaciones.

§ 5. Tabla de correlación

Para un número grande de observaciones un mismo valor de x puede encontrarse n_x veces, un mismo valor de y puede encontrarse n_y veces, un mismo par de números (x, y) puede observarse n_{xy} veces. Por eso los datos de las observaciones se agrupan, es decir, se calculan las frecuencias n_x, n_y, n_{xy}. Todos los datos agrupados se escriben en forma de tabla que se llama *de correlación.*

Aclaremos la estructura de la tabla de correlación con un ejemplo (tabla 13).

Tabla 13

Y \ X	10	20	30	40	n_y
0,4	5	—	7	14	26
0,6	—	2	6	4	12
0,8	3	19	—	—	22
n_x	8	21	13	18	$n=60$

En la primera línea de la tabla se indican los valores observados (10; 20; 30; 40) del carácter X, y en la primera columna, los valores observados (0,4; 0,6; 0,8) del carácter Y. En las intersecciones de las lineas y las columnas se escriben las frecuencias n_{xy} de los pares de valores observados de los caracteres. Por ejemplo, la frecuencia 5 indica que el par de números (10; 0,4) se observó 5 veces. Todas las frecuencias se encuentran en un rectángulo, cuyos lados se han trazado con líneas gruesas. La raya significa que el par de números correspondiente, por ejemplo (20; 0,4), no se observó.

En la última columna se escribe la suma de frecuencias de las líneas. Por ejemplo, la suma de las frecuencias de la primera línea del rectángulo de líneas gruesas es igual a $n_y = 5 + 7 + 14 = 26$; este número indica que el valor del carácter Y, igual a 0,4 (en combinación con distintos valores del carácter X) se observó 26 veces.

En la última línea se escriben la suma de frecuencias de las columnas. Por ejemplo, el número 8 indica que el valor del carácter X, igual a 10 (en combinación con distintos valores del carácter Y) se observó 8 veces.

En la casilla ubicada en el ángulo inferior derecho de la tabla, se coloca la suma de todas las frecuencias (número total de observaciones n). Evidentemente, $\sum n_x = \sum n_y = n$. En nuestro ejemplo

$\sum n_x = 8 + 21 + 13 + 18 = 60$ y

$$\sum n_y = 26 + 12 + 22 = 60.$$

§ 6. Hallazgo de los parámetros de la ecuación muestral de la recta de regresión por datos agrupados. Coeficiente de correlación muestral

En el § 4 para determinar los parámetros de la ecuación de la recta de la regresión de Y en X se obtuvo el sistema de ecuaciones:

$$\left.\begin{aligned} \left(\sum x^2\right) \rho_{yx} + \left(\sum x\right) b &= \sum xy; \\ \left(\sum x\right) \rho_{yx} + nb &= \sum y. \end{aligned}\right\} \qquad (*)$$

Se supuso que los valores de X y sus correspondientes valores de Y se observaron una vez cada uno. Ahora admitimos que se han obtenido un gran número de datos (prácticamente para una estimación satisfactoria de los parámetros buscados deben haber por lo menos 50 observaciones), entre los cuales son los que se repiten y están agrupados en forma de tabla de correlación. Escribimos el sistema (*) de manera que refleje los datos de esta tabla. Usamos las igualdades:

$$\sum x = n\bar{x} \left(\text{corolario de } \bar{x} = \frac{\sum x}{n}\right);$$

$$\sum y = n\bar{y} \left(\text{corolario de } \bar{y} = \frac{\sum y}{n}\right);$$

$$\sum x^2 = n\overline{x^2} \left(\text{corolario de } \overline{x^2} = \frac{\sum x^2}{n}\right),$$

$\sum xy = \sum n_{xy}xy$ (se ha tenido en cuenta que el par de números (x, y) se observó n_{xy} veces).

Sustituyendo los segundos miembros de las igualdades en el sistema (*) y simplificando ambos miembros de la segunda ecuación por n, obtenemos:

$$\left.\begin{aligned} (n\overline{x^2})\,\rho_{yx} + (n\overline{x})\,b &= \sum n_{xy}xy; \\ (\overline{x})\,\rho_{yx} + b &= \overline{y}. \end{aligned}\right\} \qquad (**)$$

Resolviendo este sistema, hallamos los parámetros ρ_{yx} y b, y, por lo tanto, la ecuación buscada es

$$\overline{y}_x = \rho_{yx}x + b.$$

Sin embargo, introduciendo una nueva magnitud, o sea, el coeficiente de correlación, la ecuación de regresión conviene escribirla en otra forma, lo que haremos a continuación.

Hallamos b de la segunda ecuación (**):

$$b = \overline{y} - \rho_{yx}\overline{x}.$$

Sustituyendo el segundo miembro de esta igualdad en la ecuación $\overline{y}_x = \rho_{yx}x + b$, obtenemos:

$$\overline{y}_x - \overline{y} = \rho_{yx}(x - \overline{x}). \qquad (***)$$

Del sistema (*) hallamos el coeficiente de regresión, teniendo en cuenta que $\overline{x^2} - (\overline{x})^2 = \sigma_x^2$ (cap. XVI, § 10):

$$\rho_{yx} = \frac{\sum n_{xy}xy - n\overline{x}\overline{y}}{n\,[\overline{x^2} - (\overline{x})^2]} = \frac{\sum n_{xy}xy - n\overline{x}\overline{y}}{n\sigma_x^2}.$$

Multiplicamos ambos miembros de la igualdad por $\frac{\sigma_x}{\sigma_y}$:

$$\rho_{yx}\cdot\frac{\sigma_x}{\sigma_y} = \frac{\sum n_{xy}xy - n\overline{x}\overline{y}}{n\sigma_x\sigma_y}.$$

Designamos el segundo miembro de la igualdad por r_m y lo llamamos *coeficiente de correlación muestral:*

$$\rho_{yx}\cdot\frac{\sigma_x}{\sigma_y} = r_m,$$

o bien

$$\rho_{yx}\cdot r_m\frac{\sigma_y}{\sigma_x}.$$

Poniendo el segundo miembro de esta ecuación en la (***), finalmente obtenemos la ecuación muestral de la recta de la regresión de Y en X del tipo

$$\overline{y}_x - \overline{y} = r_m \frac{\sigma_y}{\sigma_x}(x - \overline{x}).$$

Nota 1. Análogamente se halla la ecuación muestral de la línea de la regresión de X en Y de la forma

$$\overline{x}_y - \overline{x} = r_m \frac{\sigma_x}{\sigma_y}(y - \overline{y}),$$

donde

$$r_m \frac{\sigma_x}{\sigma_y} = \rho_{xy}.$$

Nota 2. Las ecuaciones de las rectas de regresión pueden ser escritas en forma más simétrica:

$$\frac{\overline{y}_x - \overline{y}}{\sigma_y} = r_m \frac{x - \overline{x}}{\sigma_x};$$

$$\frac{\overline{x}_y - \overline{x}}{\sigma_x} = r_m \frac{y - \overline{y}}{\sigma_y}.$$

Nota 3. El coeficiente de correlación muestral tiene un importante valor individual. Como se deduce de lo anterior, el coeficiente de correlación muestral se determina por la igualdad

$$r_m = \frac{\sum n_{xy} xy - n\overline{x}\overline{y}}{n\sigma_x\sigma_y},$$

donde x, y son las variantes (valores observados) de los caracteres X e Y

n_{xy} es la frecuencia del par de variantes (x, y) observadas;

n es el volumen de la muestra (suma de todas las frecuencias);

$\overline{x}$, $\overline{y}$ son las medias muestrales;

σ_x, σ_y son las desviaciones cuadráticas medias muestrales.

§ 7. Propiedades del coeficiente de correlación muestral

Fijemos las propiedades del coeficiente de correlación muestral, de las cuales se deduce que éste sirve para estimar la fuerza de la dependencia de correlación lineal.

Utilizamos las fórmulas (omitimos la deducción):

$$S_y = D_y(1 - r_m^2); \quad S_x = D_x(1 - r_m^2)$$

donde S_y es la dispersión de los valores observados de y alrededor de las correspondientes medias condicionales $\overline{y}_x$;

D_y es la dispersión de los valores observados de y alrededor de la media general $\overline{y}$.

Las dispersiones S_x, D_x tienen igual sentido.

1. *La magnitud absoluta del coeficiente de correlación muestral no es mayor que la unidad.*

DEMOSTRACION. Cualquier dispersión no es negativa. En particular,

$$S_y = D_y (1 - r_m^2) \geqslant 0.$$

Por lo tanto,

$$1 - r_m^2 \geqslant 0.$$

De aquí

$$-1 \leqslant r_m \leqslant 1,$$

o bien

$$| r_m | \leqslant 1.$$

2. *Si el coeficiente de correlación muestral es igual a cero y las líneas de regresión muestrales son rectas, tendremos que X e Y no están vinculados por una dependencia de correlación lineal.*

DEMOSTRACION. Para $r_m = 0$ la ecuación de la recta de la regresión muestral de Y en X

$$\overline{y}_x - \overline{y} = r_m \frac{\sigma_y}{\sigma_x} (x - \overline{x})$$

tiene la forma

$$\overline{y}_x - \overline{y} = 0,$$

o bien

$$\overline{y}_x = \overline{y}.$$

Si $r_m = 0$ la ecuación de la recta de la regresión de X en Y tiene la forma

$$\overline{x}_y = \overline{x}.$$

Por consiguiente, cuando $r_m = 0$ las medias condicionales conservan un valor constante al variar los correspondientes argumentos; en este sentido se puede considerar que X e Y no están vinculados por la dependencia de correlación lineal. Evidentemente, en el caso examinado las rectas de regresión son paralelas a los correspondientes ejes de coordenadas.

Nota. Si el coeficiente de correlación muestral es igual a cero, los caracteres X e Y pueden estar vinculados por una dependencia de correlación *no lineal* e incluso funcional.

3. *Si la magnitud absoluta del coeficiente de correlación muestral es igual a la unidad, losv alores observados de los caracteres están vinculados por una dependencia funcional lineal.*

$$\text{Si } |r_m| = 1, \text{ entonces } S_y = D_y(1 - r_m^2) = 0.$$

Se puede mostrar que de aquí se deduce la igualdad:

$$y - \bar{y} - r_m \frac{\sigma_y}{\sigma_x}(x - \bar{x}) = 0.$$

Como podemos apreciar, todo par de números observado (x, y) satisface esta ecuación *lineal* respecto a x e y, es decir, los valores de los caracteres de la muestra están vinculados por una dependencia funcional lineal. Cabe hacer notar que de aquí no podemos deducir con certeza que también en el conjunto general los caracteres están vinculados por una dependencia funcional lineal (para una muestra representativa de gran volumen la dependencia entre los caracteres de un conjunto general normalmente distribuido será próxima a la lineal, o incluso será lineal).

4. *Al aumentar el valor absoluto del coeficiente de correlación muestral la dependencia de correlación lineal deviene más estrecha y para* $|r_m| = 1$ *pasa a la dependencia funcional.*

DEMOSTRACION. De las fórmulas

$$S_y = D_y(1 - r_m^2), \quad S_x = D_x(1 - r_m^2)$$

se aprecia que al crecer el valor absoluto de r_m las dispersiones S_y y S_x decrecen, es decir, disminuye la dispersión de los valores observados de los caracteres alrededor de las medias condicionales, y esto, precisamente, significa que el vínculo entre los caracteres se hace más estrecho y cuando $|r_m| = 1$, como se desprende de la propiedad 3, pasa a la funcional.

De las propiedades expuestas se desprende el sentido de r_m: *el coeficiente de correlación muestral caracteriza la estrechez del vínculo* lineal *entre los caracteres cuantitativos de la muestra; cuanto más próximo* $|r_m|$ *es a* 1, *tanto más fuerte es el enlace; cuanto más próximo* $|r_m|$ *es a* 0, *tanto más débil es el enlace.*

Si la muestra tiene un volumen suficientemente grande y representa bien el conjunto general (es representativa), la conclusión sobre la estrechez de la dependencia lineal entre caracteres, obtenida por los datos de la muestra, en cierto grado puede ser extendida al *conjunto general*. Por ejemplo, para estimar el coeficiente de correlación r_g de un conjunto general normalmente distribuido (para $n \geqslant 50$) se puede utilizar la fórmula

$$r_m - 3\frac{1-r_m^2}{\sqrt{n}} \leqslant r_g \leqslant r_m + 3\frac{1+r_m^2}{\sqrt{n}}.$$

Nota 1. El signo del coeficiente de correlación muestral coincide con el signo de los coeficientes de regresión muestrales, lo que se desprende de las fórmulas (§ 4):

$$\rho_{yx} = r_m \frac{\sigma_y}{\sigma_x}; \quad \rho_{xy} = r_m \frac{\sigma_x}{\sigma_y}. \qquad (*)$$

Nota 2. El coeficiente de correlación muestral es igual a la media geométrica de los coeficientes muestrales de regresión. En efecto, multiplicando miembro a miembro, primeros y segundos, de las igualdades (*) obtenemos:

$$\rho_{yx}\rho_{xy} = r_m^2.$$

De aquí

$$r_m = \pm\sqrt{\rho_{yx}\rho_{xy}}.$$

El signo que afecta el radical, de acuerdo con la nota 1, debe coincidir con el signo de los coeficientes de regresión.

§ 8. Método de los cuatro campos para el cálculo del coeficiente de correlación muestral

Supongamos que por los datos de la tabla de correlación hay que calcular el coeficiente de correlación muestral. El cálculo se simplifica considerablemente si se pasa a las variantes condicionales:

$$u_i = \frac{x_i - c_1}{h_1} \text{ y } v_j = \frac{y_j - c_2}{h_2}.$$

En este caso el coeficiente de correlación muestral se calcula por la fórmula (la transición a las variantes condicionales no altera la magnitud r_m):

$$r_m = \frac{\sum n_{uv}uv - n\bar{u}\bar{v}}{n\sigma_u\sigma_v}.$$

Las magnitudes $\overline{u}$, $\overline{v}$, σ_u y σ_v pueden ser calculadas por el método de los productos (cap. XVII, § 4). Queda por enseñar el método de cálculo de $\sum n_{uv}uv$. Precisamente para esto se utiliza el *método de los cuatro campos*. El nombre de este método se debe a que la línea y la columna que se intersecan en la casilla que contiene la frecuencia máxima, dividen la tabla de correlación en 4 partes, llamadas *campos*. Los campos se enumeran como está indicado en la tabla 14.

Tabla 14

$v \backslash u$		0	
	I		II
0		frecuen. máxima	
	III		IV

Vamos a mostrar cómo se realiza el cálculo, para lo cual nos limitaremos por ahora al campo I. Supongamos que la parte de la tabla que contiene el primer campo, está representada en la forma de la tabla 15.

Tabla 15

$v \backslash u$	−3	−2	−1
−2	5	1	—
−1	—	20	23

Hallamos los productos de los pares de variantes u y v, los alojamos en los ángulos superiores derechos de las casillas que contienen las frecuencias correspondientes. Por ejemplo, el par de variantes $u = -3$ y $v = -2$ se observó 5 veces; el

producto $uv = (-3)\cdot(-2) = 6$ lo colocamos en el ángulo superior derecho de la casilla que contiene la frecuencia 5. Llenando de manera semejante las restantes casillas del primer campo, obtenemos la tabla 16.

Tabla 16

v \ u	−3	−2	−1
−2	6 / 5	4 / 7	—
−1	—	2 / 20	1 / 23

Análogamente se llenan las casillas de los demás campos. Por consiguiente, en cada casilla (que contiene la frecuencia n_{uv}), resulta también escrito el producto uv, queda multiplicar los dos números n_{uv} y uv de cada casilla y sumar los resultados; en conclusión obtenemos el número buscado $\sum n_{uv}uv$.

Para verificar convenientemente los cálculos los productos de los números n_{uv} y uv hallados de cada casilla se suman separadamente por cada campo; además, se calcula también por líneas y por columnas de cada campo. La suma de los números $n_{uv}\cdot uv$ de las líneas del campo se escribe en aquella de las columnas adicionales dispuestas a la derecha que tiene la cifra del campo, cuyos números se sumaron. La suma de los números $n_{uv}\cdot uv$ de las columnas del campo se escribe en aquella de las líneas adicionales ubicadas abajo que tiene la cifra del campo, cuyos números se sumaron. Las sumas de los números de cada campo por separado se escriben en el ángulo inferior derecho de la tabla en cuatro casillas de totales. Por último, sumando todos los números de las casillas de totales, se halla el número buscado.

Esquemáticamente la tabla de cálculo tiene la forma de la tabla 17.

Expliquemos cómo se ha completado la tabla 17 (para mayor claridad realizaremos el cálculo sólo para el primer campo).

v \ u	−3	−2	−1	0		I	II
−2	5 (6)	7 (4)	—			58	
−1	—	20 (2)	23 (1)		II	63	
0				frecuen. máxima		III	IV
		III			IV		
I	30	68	23	II		121	II
III				IV		III	IV

Hallamos las sumas de los productos n_{uv} y uv de las líneas del primer campo ($5 \cdot 6 + 7 \cdot 4 = 58$; $20 \cdot 2 + 23 \cdot 1 = 63$) y las colocamos en la columna adicional I.

Hallamos las sumas de los productos n_{uv} y uv de las columnas del primer campo ($5 \cdot 6 = 30$; $7 \cdot 4 + 20 \cdot 2 = 68$; $23 \cdot 1 = 23$) y las colocamos en la línea adicional I.

Hallamos la suma de los números de la columna adicional I ($58 + 63 = 121$) y la escribimos en la primera casilla de total (en el primer ángulo inferior de la tabla).

Para verificar sumamos todos los números de la línea adicional ($30 + 68 + 23 = 121$).

Análogamente se calculan por los demás campos.

Ejemplo. Calcular el coeficiente de correlación muestral por los datos de las tabla de correlación 18.

SOLUCIÓN. Pasamos a las variantes condicionales: $u = \frac{x - c_1}{h_1} = \frac{x - 40}{10}$ (como cero accidental c_1 se ha tomado la variante $x = 40$ que tiene la frecuencia máxima; el

paso h_1 es igual a la diferencia entre dos variantes contiguas: $20 - 10 = 10$) y

Tabla 18

Y \ X	10	20	30	40	50	60	n_y
15	5	7	—	—	—	—	12
25	—	20	23	—	—	—	43
35	—	—	30	47	2	—	79
45	—	—	10	11	20	6	47
55	—	—	—	9	7	3	19
n_x	5	27	63	67	29	9	$n=200$

$v=\frac{y-c_2}{h_2}=\frac{y-35}{10}$ (como cero accidental c_2 se ha tomado la variante $y=35$ que tiene la frecuencia máxima; el paso h_2 es igual a la diferencia entre dos variantes contiguas $25-15=10$).

Formamos la tabla de correlación en las variantes condicionales. Prácticamente se procede del modo siguiente: en la primera columna en vez de la variante (35), de frecuencia máxima, se escribe el 0; sobre el cero se escribe sucesivamente -1, -2; bajo el cero se escribe 1, 2. En la primera línea en lugar de la variante (40), que tiene la frecuencia máxima, se escribe el 0; a la izquierda del cero se escribe sucesivamente -1, -2, -3; a la derecha del cero se escribe 1, 2. Todos los demás datos se copian de la tabla de correlación original. Como resultado obtenemos la tabla de correlación 19 en variantes condicionales.

Las magnitudes $\bar{u}$, $\bar{v}$, σ_u y σ_v se pueden hallar por el método de los productos; sin embargo, puesto que los números u_i, v_j son pequeños, calculamos $\bar{u}$ y $\bar{v}$, partiendo de la

definición del valor medio, y σ_u y σ_v, utilizando las fórmulas (cap. XVI, § 10):

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2}, \quad \sigma_v = \sqrt{\overline{v^2} - (\bar{v})^2}.$$

Tabla 19

v \ u	−3	−2	−1	0	1	2	n_v
−2	5	7	—	—	—	—	12
−1	—	20	23	—	—	—	43
0	—	—	30	47	2	—	79
1	—	—	10	11	20	6	47
2	—	—	—	9	7	3	19
n_u	5	27	63	67	29	9	$n=200$

Hallamos $\bar{u}$ y $\bar{v}$:

$$\bar{u} = \frac{\sum n_u u}{n} = \frac{5\cdot(-3)+27\cdot(-1)+63\cdot(-2)+29\cdot 1+9\cdot 2}{200} = -0{,}425;$$

$$\bar{v} = \frac{\sum n_v v}{n} = \frac{12\cdot(-2)+43\cdot(-1)+47\cdot 1+19\cdot 2}{200} = 0{,}09.$$

Calculamos la magnitud auxiliar $\overline{u^2}$, y luego σ_u:

$$\overline{u^2} = \frac{5\cdot 9+27\cdot 4+63\cdot 1+29\cdot 1+9\cdot 4}{200} = 1{,}405;$$

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{1{,}405 - 0{,}425^2} = 1{,}106.$$

Análogamente obtenemos $\sigma_v = 1{,}209$.

Hallamos $\sum n_{uv} uv$ por el método de los cuatro campos, para lo cual formamos la tabla de cálculo 20.

Tabla 20

	−3	−2	−1	0	1	2	I	II
−2	5 [6]	7 [4]	—		—	—	58	—
−1	—	20 [2]	23 [1]		—	—	63	—
0							III	IV
1	—	—	10 [−1]		20 [1]	6 [2]	−10	32
2	—	—	—		7 [2]	3 [4]	—	26
I	30	68	23	II	—	—	121	—
III	—	—4	−10	IV	34	24	−10	58

Sumando los números de las casillas de total (4 casillas en el ángulo inferior derecho de la tabla 20), obtenemos

$$\sum n_{uv}uv = 121 - 10 + 58 = 169.$$

Hallamos el coeficiente de correlación buscado:

$$r_m = \frac{\sum n_{uv}uv - n\bar{u}\bar{v}}{n\sigma_u\sigma_v} = \frac{169 - 200\cdot(-0{,}425)\cdot 0{,}09}{200\cdot 1{,}106\cdot 1{,}209} = 0{,}603.$$

De este modo,

$$r_m = 0{,}603.$$

§ 9. Ejemplo de hallazgo de la ecuación de la recta de regresión muestral

Ahora, cuando sabemos como calcular r_m, es oportuno dar un ejemplo de cómo hallar la ecuación de la recta de regresión.

Puesto que para hallar r_m ya se han calculado $\overline{u}$, $\overline{v}$, σ_u, σ_v, conviene utilizar las fórmulas:

$$\sigma_x = h_1\sigma_u, \quad \sigma_y = h_2\sigma_v, \quad \overline{x} = \overline{u}h_1 + c_1,$$
$$\overline{y} = \overline{v}h_2 + c_2.$$

Aquí se han conservado las notaciones del párrafo anterior. Recomendamos a los lectores deducir individualmente estas fórmulas.

Ejemplo. Hallar la ecuación muestral de la recta de la regresión de Y en X por los datos de la tabla de correlación 18 del ejemplo del párrafo precedente.

SOLUCION. Escribimos la ecuación buscada en forma general:

$$\overline{y}_x - \overline{y} = r_m \frac{\sigma_y}{\sigma_x}(x - \overline{x}). \qquad (*)$$

El coeficiente de correlación se calculó en el párrafo anterior. Queda hallar $\overline{x}$, $\overline{y}$, σ_x y σ_y:

$$\overline{x} = \overline{u}h_1 + c_1 = -0{,}425 \cdot 10 + 40 = 35{,}75;$$
$$\overline{y} = \overline{v}h_2 + c_2 = 0{,}09 \cdot 10 + 35 = 35{,}9;$$
$$\sigma_x = \sigma_u h_1 = 1{,}106 \cdot 10 = 11{,}06;$$
$$\sigma_y = \sigma_v h_2 = 1{,}209 \cdot 10 = 12{,}09.$$

Sustituyendo las magnitudes halladas en (*), obtenemos la ecuación buscada

$$\overline{y}_x - 35{,}9 = 0{,}603\, \frac{12{,}09}{11{,}06}(x - 35{,}75),$$

o bien, finalmente

$$\overline{y}_x = 0{,}659x + 12{,}34.$$

Comparamos las medias condicionales calculadas: a) por esta ecuación; b) por los datos de la tabla de correlación 18. Por ejemplo, para $x = 30$:

$$\text{a)}\ \overline{y}_{30} = 0{,}659 \cdot 30 + 12{,}34 = 32{,}11;$$
$$\text{b)}\ \overline{y}_{30} = \frac{23 \cdot 25 + 30 \cdot 35 + 10 \cdot 45}{63} = 32{,}94.$$

Como podemos apreciar, la concordancia de las medias condicionales calculada y observada es satisfactoria.

§ 10. Consideraciones preliminares al establecimiento de la medida de cualquier enlace de correlación

Ya hemos examinado la estimación de la fuerza de un enlace de correlación *lineal*. ¿Cómo estimar la fuerza de *cualquier* enlace de correlación?

Supongamos que los datos de las observaciones de los caracteres cuantitativos X e Y están expuestos en una tabla de correlación. Podemos considerar que los mismos valores observados de Y están descompuestos en grupos; cada grupo contiene los valores de Y que corresponden a un valor determinado de X.

Por ejemplo, está dada la tabla de correlación 21.

Tabla 21

Y \ X	8	9
3	4	13
5	6	7
n_x	10	20
$\bar{y}_x$	4,2	3,7

Al primer grupo pertenecen los 10 valores de Y (4 veces se observó $y_1 = 3$ y 6 veces $y_2 = 5$) que corresponden a $x_1 = 8$.

Al segundo grupo pertenecen los 20 valores de Y (13 veces se observó $y_1 = 3$ y 7 veces $y_2 = 5$) que corresponden a $x_2 = 9$.

Las medias condicionales ahora las podemos llamar medias de grupos: la media de grupo del primer grupo es $\bar{y}_8 = \frac{4 \cdot 3 + 6 \cdot 5}{10} = 4{,}2$; la media de grupo del segundo grupo es $\bar{y}_9 = \frac{13 \cdot 3 + 7 \cdot 5}{20} = 3{,}7$.

Puesto que todos los valores del carácter Y están descompuestos en grupos, la dispersión general del carácter se puede

representar en forma de suma de dispersiones dentro de grupo y entre grupo (cap. XVI, § 12):

$$D_{\text{gen}} = D_{\text{den. gru}} + D_{\text{ent. gru}}. \qquad (*)$$

Demostraremos el cumplimiento de las tesis siguientes:

1) si Y está vinculada con X por una dependencia funcional, tendremos que

$$\frac{D_{\text{ent. gru}}}{D_{\text{gen}}} = 1;$$

2) si Y está vinculada con X por una dependencia de correlación, entonces

$$\frac{D_{\text{ent. gru}}}{D_{\text{gen}}} < 1.$$

DEMOSTRACION. 1) Si Y está vinculada con X *por una dependencia funcional*, a un valor determinado de X corresponde un solo valor de Y. En este caso, cada grupo contiene valores de Y* iguales entre sí, por eso la dispersión de grupo de cada grupo es igual a cero. Por lo tanto, la media aritmética de las dispersiones de grupos (ponderada por los volúmenes de los grupos), es decir, la dispersión dentro de grupo $D_{\text{den gru}} = 0$ y la igualdad (*) tiene la forma

$$D_{\text{gen}} = D_{\text{ent gru}}.$$

De aquí

$$\frac{D_{\text{ent. gru}}}{D_{\text{gen}}} = 1.$$

2) Si Y está vinculada con X por una *dependencia de correlación*, a un valor determinado de X corresponden, en general, distintos valores de Y (que forman el grupo). En este caso la dispersión de grupo de cada grupo es distinta de cero. Por lo tanto, la media aritmética de las dispersiones de grupos (ponderada por los volúmenes de los grupos) $D_{\text{den gru}} \neq 0$. Entonces (un sumando positivo, es decir $D_{\text{ent gru}}$ es menor que la suma de dos sumandos positivos $D_{\text{den gru}} + D_{\text{ent gru}} = D_{\text{gen}}$):

$$D_{\text{ent. gru}} < D_{\text{gen}}.$$

De aquí

$$\frac{D_{\text{ent. gru}}}{D_{\text{gen}}} < 1.$$

---

* Por ejemplo, si al valor de $x_1 = 3$ corresponde $y_1 = 7$; además, $x_1 = 3$ se observó 5 veces, el grupo contiene 5 valores de $y_1 = 7$.

De los razonamientos expuestos se aprecia que cuanto más próximo a la funcional es el vínculo entre los caracteres, tanto menor es $D_{den.gru}$, y, por lo tanto, tanto más se aproximará $D_{ent.gru}$ a D_{gen}, y en consecuencia, la relación $\frac{D_{ent.gru}}{D_{gen}}$ a la unidad. De aquí está claro que como medida de la fuerza de la dependencia de correlación conviene considerar la relación entre la dispersión entre grupo y la general o bien, lo que es equivalente, la relación entre la desviación cuadrática media entre grupo y la desviación cuadrática media general.

§ 11. Relación de correlación muestral

Para estimar la fuerza del enlace de correlación *lineal* entre los caracteres de una muestra se utiliza el coeficiente de correlación muestral. Para estimar la fuerza del enlace de correlación *no lineal* se introducen nuevas características generales:

η_{yx}, relación de correlación muestral entre Y y X;

η_{xy}, relación de correlación muestral entre X e Y.

Se llama *relación de correlación muestral* de Y a X la relación entre la desviación cuadrática media entre grupos y la desviación cuadrática media general del carácter Y

$$\eta_{yx}=\frac{\sigma_{ent.\,gru}}{\sigma_{gen}},$$

o bien en otras notaciones

$$\eta_{yx}=\frac{\sigma_{\overline{y}_x}}{\sigma_y}.$$

Aquí

$$\sigma_{\overline{y}_x}=\sqrt{D_{ent.\,gru}}=\sqrt{\frac{\sum n_x(\overline{y}_x-\overline{y})^2}{n}};$$

$$\sigma_y=\sqrt{D_{gen}}=\sqrt{\frac{\sum n_y(y-\overline{y})^2}{n}},$$

donde n es el volumen de la muestra (suma de todas las frecuencias);

n_x es la frecuencia del valor x del carácter X;

n_y es la frecuencia del valor y del carácter Y;

$\overline{y}$ es la media general del carácter Y;

$\overline{y}_x$ es la media condicional del carácter Y

Análogamente se determina la relación de correlación muestral de X a Y:

$$\eta_{xy} = \frac{\sigma_{\bar{x}_y}}{\sigma_x}.$$

Ejemplo. Hallar η_{yx} por los datos de la tabla de correlación 22.

Tabla 22

Y \ X	10	20	30	n_y
15	4	28	6	38
25	6	—	6	12
n_x	10	28	12	$n=50$
$\bar{y}_x$	21	15	20	

SOLUCION. Hallamos la media general

$$\bar{y} = \frac{\sum n_y y}{n} = \frac{38 \cdot 15 + 12 \cdot 25}{50} = 17,4.$$

Encontramos la desviación cuadrática media general

$$\sigma_y = \sqrt{\frac{\sum n_y (y - \bar{y})^2}{n}} =$$

$$= \sqrt{\frac{38\,(15 - 17,4)^2 + 12\,(25 - 17,4)^2}{50}} = 4,27.$$

Hallamos la desviación cuadrática media entre grupo

$$\sigma_{\bar{y}_x} = \sqrt{\frac{\sum n_x (\bar{y}_x - \bar{y})^2}{n}} =$$

$$= \sqrt{\frac{10\,(21 - 17,4)^2 + 28\,(15 - 17,4)^2 + 12\,(20 - 17,4)^2}{50}} = 2,73.$$

La relación de correlación buscada es

$$\eta_{yx} = \frac{\sigma_{\bar{y}_x}}{\sigma_y} = \frac{2,73}{4,27} = 0,64.$$

§ 12. Propiedades de la relación de correlación muestral

Puesto que η_{xy} tiene las mismas propiedades que η_{yx} enumeraremos solamente las propiedades de la relación de correlación muestral η_{yx} cuya notación, para simpleza, en adelante la designaremos por η y, para sencillez verbal, la llamaremos «relación de correlación».

1. *La relación de correlación satisface la doble desigualdad*

$$0 \leqslant \eta \leqslant 1.$$

DEMOSTRACION. La desigualdad $\eta \geqslant 0$ resulta de que η es una relación de números no negativos, o sea, de desviaciones cuadráticas medias (de entre grupos al general).

Para demostrar la desigualdad $\eta \leqslant 1$ utilizamos la fórmula

$$D_{\text{gen}} = D_{\text{den.gru}} + D_{\text{ent.gru}}.$$

Dividiendo ambos miembros de la igualdad por D_{gen}, obtenemos:

$$1 = \frac{D_{\text{den. gru}}}{D_{\text{gen}}} + \frac{D_{\text{ent. gru}}}{D_{\text{gen}}},$$

o bien

$$1 = \frac{D_{\text{den. gru}}}{D_{\text{gen}}} + \eta^2.$$

Puesto que ambos sumandos no son negativos y su suma es igual a la unidad, cada uno de ellos no es mayor que la unidad; en particular

$$\eta^2 \leqslant 1.$$

Teniendo en cuenta que $\eta \geqslant 0$, deducimos:

$$0 \leqslant \eta \leqslant 1.$$

2. *Si* $\eta = 0$, *tendremos que el carácter Y no está vinculado con el carácter X por una dependencia de correlación.*

DEMOSTRACION. Por los datos

$$\eta = \frac{\sigma_{\text{ent. gru}}}{\sigma_{\text{gen}}} = 0.$$

De aquí

$$\sigma_{\text{ent.gru}} = 0,$$

y, por lo tanto,

$$D_{\text{ent.gru}} = 0.$$

La dispersión entre grupos es la dispersión de las medias condicionales (de grupo) $\overline{y}_x$ con respecto a la media general $\overline{y}$. La dispersión entre grupos igual a cero significa que para todos los valores de X las medias condicionales conservan un valor constante (igual a la media general). En otras palabras, cuando $\eta = 0$ la media condicional no es una función de X, lo que significa que el carácter Y no está vinculado por una dependencia de correlación con el carácter X.

Nota 1. Se puede demostrar también la enunciación inversa: si el carácter Y no está vinculado con el carácter X por una dependencia de correlación, tendremos que $\eta = 0$.

3. *Si* $\eta = 1$, *el carácter* Y *está vinculado con el carácter* X *por una dependencia funcional.*

DEMOSTRACIÓN. Por los datos

$$\eta = \frac{\sigma_{\text{ent. gru}}}{\sigma_{\text{gen}}} = 1.$$

De aquí

$$\sigma_{\text{gen}} = \sigma_{\text{ent. gru}}.$$

Elevando al cuadrado ambos miembros de la igualdad, obtenemos

$$D_{\text{gen}} = D_{\text{ent. gru}}. \qquad (*)$$

Puesto que

$$D_{\text{gen}} = D_{\text{den. gru}} + D_{\text{ent. gru}},$$

en virtud de (*)

$$D_{\text{den. gru}} = 0. \qquad (**)$$

Dado que la dispersión dentro de grupo es la media aritmética de las dispersiones de grupos (ponderadas por los volúmenes de los grupos), de la (**) se deduce que la dispersión de cada grupo (de los valores de Y correspondientes a un valor determinado de X) es igual a cero. Esto significa que el grupo tiene iguales valores de Y, es decir, a cada valor de X corresponde un valor de Y. Por lo tanto, cuando $\eta = 1$ el carácter Y está vinculado con el carácter X por una dependencia funcional.

Nota 2. También se puede demostrar la enunciación inversa: si el carácter Y está vinculado con el carácter X por una dependencia funcional, entonces $\eta = 1$.

Enunciamos dos propiedades más sin demostración.

4. *La relación de correlación muestral no es menor que la magnitud absoluta del coeficiente de correlación muestral*

$$\eta \geqslant |r_m|.$$

5. *Si la relación de correlación muestral es igual a la magnitud absoluta del coeficiente de correlación muestral, tiene lugar una dependencia de correlación lineal exacta.*

En otras palabras, si $\eta = |r_m|$, los puntos

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

se encuentran sobre la recta de regreción hallada por el método de cuadrados mínimos.

§ 13. Relación de correlación como medida de enlace de correlación. Méritos e insuficiencias de esta medida

En el párrafo anterior se estableció que: si $\eta = 0$, los caracteres no están vinculados por una dependencia de correlación; si $\eta = 1$, tiene lugar la dependencia funcional.

Demostremos que al crecer η el enlace de correlación se hace más estrecho. Para ello transformemos la correlación

$$D_{gen} = D_{den\ gru} + D_{ent.gru}$$

del modo siguiente:

$$D_{den.\,gru} = D_{gen}\left(1 - \frac{D_{ent.\,gru}}{D_{gen}}\right),$$

o bien

$$D_{den\ gru} = D_{gen}\,(1 - \eta^2).$$

Si $\eta \to 1$, entonces $D_{den.gru} \to 0$, por lo tanto, también tiende a cero cada una de las dispersiones de grupos. En otras palabras, al aumentar η los valores de Y correspondientes a un valor determinado de X, se diferencian cada vez menos entre sí y el vínculo de Y con X deviene más estrecho, pasando a la funcional, cuando $\eta = 1$.

Ya que en los razonamientos no se hicieron suposiciones sobre la forma del enlace de correlación, η *sirve de medida de la estrechez del enlace de cualquier forma, incluso lineal.* En esto reside la ventaja de la relación de correlación en

comparación con el coeficiente de correlación que estima solamente la estrechez de la dependencia lineal. Al mismo tiempo, la relación de correlación tiene *insuficiencia*: no permite juzgar cuán próximos se encuentran los puntos hallados por los datos de las observaciones, a una curva de forma determinada, por ejemplo, a una parábola, hipérbola, etc. Esto se debe a que no tomamos en consideración la forma del enlace al definir la relación de correlación.

§ 14. Casos elementales de correlación curvilínea

Si la gráfica de la regresión $\overline{y}_x = f(x)$ o bien $\overline{x}_y = \varphi(y)$ se representa por una curva, la correlación se llama *curvilínea*.

Por ejemplo, la función de la regresión de Y en X puede tener la forma:

$\overline{y}_x = ax^2 + bx + c$ (correlación parabólica de segundo orden);

$\overline{y}_x = ax^3 + bx^2 + cx + d$ (correlación parabólica de tercer orden);

$\overline{y}_x = \frac{a}{x} + b$ (correlación hiperbólica).

La teoría de la correlación curvilínea resuelve los mismos problemas que la teoría de la correlación lineal (establecimiento de la forma y la estrechez del enlace de correlación).

Los parámetros desconocidos de la ecuación de regresión se buscan por el método de cuadrados mínimos. Para estimar la estrechez de la correlación curvilínea se utilizan las relaciones de correlación muestrales (§ 11).

Para aclarar la esencia del problema, nos limitaremos a la correlación parabólica de segundo orden, suponiendo que los datos de n observaciones (muestra) permiten considerar que precisamente tiene lugar esta correlación. En este caso, la ecuación de la regresión muestral de Y en X tiene la forma:

$$\overline{y}_x = Ax^2 + Bx + C, \qquad (*)$$

donde A, B, C son parámetros desconocidos.

Aplicando el método de cuadrados mínimos se obtiene el sistema de ecuaciones lineales respecto a los parámetros incógnitas (omitimos la deducción, puesto que no contiene

nada nuevo en comparación con el § 4):

$$\left.\begin{aligned}\left(\sum n_x x^4\right) A + \left(\sum n_x x^3\right) B + \left(\sum n_x x^2\right) C &= \sum n_x \overline{y}_x x^2;\\ \left(\sum n_x x^3\right) A + \left(\sum n_x x^2\right) B + \left(\sum n_x x\right) C &= \sum n_x \overline{y}_x x;\\ \left(\sum n_x x^2\right) A + \left(\sum n_x x\right) B + nC &= \sum n_x \overline{y}_x.\end{aligned}\right\} \quad (**)$$

Los parámetros A, B, C hallados de este sistema se sustituyen en (*), obteniendo como resultado la ecuación de regresión buscada.

Ejemplo. Hallar la ecuación de la regresión muestral de Y en X del tipo $\overline{y}_x = Ax^2 + Bx + C$ por los datos de la tabla de correlación 23.

Formamos la tabla de cálculo 24.

Poniendo los números (sumas) de la línea inferior de la tabla 24 en (**) obtenemos el sistema:

$$\left.\begin{aligned}74{,}98\ A + 67{,}48\ B + 60{,}89\ C &= 413{,}93,\\ 67{,}48\ A + 60{,}89\ B + 55{,}10\ C &= 373{,}30,\\ 60{,}89\ A + 55{,}10\ B + 50\ C &= 337{,}59.\end{aligned}\right\}$$

Tabla 23

Y \ X	1	1,1	1,2	n_y
6	8	2	—	10
7	—	30	—	30
7,5	—	1	9	10
n_x	8	33	9	$n = 50$
$\overline{y}_x$	6	6,73	7,5	

Tabla 24

x	n_x	$\bar{y}_x$	$n_x x$	$n_x x^2$	$n_x x^3$	$n_x x^4$	$n_x \bar{y}_x$	$n_x \bar{y}_x x$	$n_x \bar{y}_x x^2$
1	8	6	8	8	8	8	48	48	48
1,1	33	6,73	36,3	39,93	43,93	48,32	222,09	244,30	268,73
1,2	9	7,5	10,8	12,96	15,55	18,66	67,50	81	97,20
$\sum$	50	—	55,1	60,89	67,48	74,98	337,59	373,30	413,93

Resolviendo este sistema, hallamos

$$A = 1{,}94, \; B = 2{,}98, \; C = 1{,}10.$$

Escribimos la ecuación de regresión buscada:

$$\bar{y}_x = 1{,}94x^2 + 2{,}98x + 1{,}10.$$

Se demuestra fácilmente que las medias condicionales, calculadas por esta ecuación, se diferencian poco de las medias condicionales de la tabla de correlación. Por ejemplo, para $x_1 = 1$ hallamos: por la tabla $\bar{y}_1 = 6$; por la ecuación $\bar{y}_1 = 1{,}94 + 2{,}98 + 1{,}10 = 6{,}02$. Por consiguiente, la ecuación hallada concuerda bien con los datos de las observaciones (muestra).

§ 15. Concepto de correlación múltiple

Hasta aquí se examinó el enlace de correlación entre dos caracteres. Si se estudia el enlace (o vínculo) entre varios caracteres, la correlación se llama *múltiple*.

En el caso elemental el número de caracteres es igual a tres, y el enlace entre ellos, lineal:

$$z = ax + by + c.$$

En este caso surgen los siguientes problemas:

1) hallar por los datos de las observaciones la ecuación de enlace del tipo

$$z = Ax + By + C, \qquad (*)$$

es decir, hay que encontrar los coeficientes de regresión de A y B y el parámetro C;

2) estimar la estrechez de enlace entre Z y ambos caracteres X, Y;

3) estimar la estrechez de enlace entre Z y X (para Y constante), entre Z e Y (para X constante).

El primer problema se resuelve por el método de cuadrados mínimos; además, en lugar de la ecuación (*) conviene buscar una ecuación de enlace de tipo

$$z-\bar{z}=A(x-\bar{x})+B(y-\bar{y}),$$

donde

$$A=\frac{r_{xz}-r_{yz}r_{xy}}{1-r_{xy}^2}\cdot\frac{\sigma_z}{\sigma_x};\quad B=\frac{r_{yz}-r_{xz}r_{xy}}{1-r_{xy}^2}\cdot\frac{\sigma_z}{\sigma_y}.$$

Aquí r_{xz}, r_{yz}, r_{xy} son coeficientes de correlación respectivamente entre los caracteres X y Z, Y y Z, X e Y; σ_x, σ_y, σ_z son las desiviaciones cuadráticas medias.

La estrechez de enlace del carácter Z con los caracteres X e Y se estima por el *coeficiente de correlación común muestral*:

$$R=\sqrt{\frac{r_{xz}^2-2r_{xy}r_{xz}r_{yz}+r_{yz}^2}{1-r_{xy}^2}},$$

además $0\leqslant R\leqslant 1$.

La estrechez de enlace entre Z y X (para Y constante), entre Z e Y (para X constante) se estima respectivamente *por los coeficientes de correlación parciales muestrales*:

$$r_{xz(y)}=\frac{r_{xz}-r_{xy}r_{yz}}{\sqrt{(1-r_{xy}^2)(1-r_{yz}^2)}};$$

$$r_{yz(x)}=\frac{r_{yz}-r_{xy}r_{xz}}{\sqrt{(1-r_{xy}^2)(1-r_{xz}^2)}}.$$

Estos coeficientes tienen las mismas propiedades y el mismo sentido que el coeficiente de correlación muestral ordinario, es decir, sirven para estimar el enlace lineal entre caracteres.

Problemas

En los problemas 1—2 se dan las tablas de correlación. Hallar: a) r_m; b) las ecuaciones de las rectas de regresión muestrales; c) η_{yx} y η_{xy}.

1.

Y \ X	5	10	15	20	n_y	$\bar{x}_y$
10	2	—	—	—	2	5
20	5	4	1	—	10	8
30	3	8	6	3	20	12,25
40	—	3	6	6	15	16
50	—	—	2	1	3	16,67
n_x	10	15	15	10	$n=50$	
$\bar{y}_x$	21	29,33	36	38		

Respuesta a) 0,636;

b) $\bar{y}_x = 1,17x + 16,78$; $\bar{x}_y = 0,345y + 1,67$;

c) $\eta_{yx} = 0,656$, $\eta_{xy} = 0,651$.

2.

Y \ X	65	95	125	155	185	215	n_y	$\bar{x}_y$
30	5		—	—	—	—	5	65
40	4	12	—	—	—	—	16	87,5

Continuación de la tabla 2)

Y \ X	65	95	125	155	185	215	n_y	$\overline{x}_y$
50	—	8	5	4	—	—	17	101,18
60	—	1	5	7	2	—	15	145
70	—	—	—	—	1	1	2	200
n_x	9	21	10	11	3	1	$n=55$	
$\overline{y}_x$	34,44	44,76	55	56,36	63,33	70		

Respuesta a) 0,825;

$$\overline{y}_x = 0{,}23x + 21{,}78,\quad \overline{x}_y = 2{,}92\, y - 27{,}25$$
$$\eta_{yx} = 0{,}859,\quad \eta_{xy} = 0{,}875.$$

En los problemas 3—4 hay que hallar las ecuaciones de regresión muestrales $\overline{y}_x = Ax^2 + Bx + C$ por los datos de la tabla de correlación.

3.

Y \ X	2	3	5	n_y
25	20	—	—	20
45	—	30	1	31
110	—	1	48	49
	20	31	49	$n=100$

Respuesta $\overline{y}_x = 2{,}92x^2 + 7{,}27x - 1{,}25.$

4.

Y \ X	1	2	n_y
2	30	1	31
6	1	18	19
n_x	31	19	$n=50$

Respuesta $\overline{y}_x = 0{,}39x^2 + 2{,}49x - 0{,}75.$

Capítulo diez y nueve

VERIFICACION ESTADISTICA DE LAS HIPOTESIS ESTADISTICAS

§ 1. Hipótesis estadística. Hipótesis nula y concurrente, simple y compleja

Frecuentemente hay que conocer la ley de distribución de un conjunto general. Si la ley de distribución es desconocida, pero existen las bases para suponer que ella tiene una forma determinada (llamémosla A), presentamos la hipótesis: el conjunto general está distribuido por la ley A. Por consiguiente, en esta hipótesis se habla *del tipo de distribución supuesta.*

Puede darse el caso en que se conoce la ley de distribución, pero no sus parámetros. Si hay razón de suponer que el parámetro incógnita Θ es igual a un valor determinado Θ_0, formulamos la hipótesis: $\Theta = \Theta_0$. Por consiguiente, en esta hipótesis se habla *de la magnitud supuesta del parámetro* de una distribución conocida.

Son posibles, también, otras hipótesis: sobre la igualdad de los parámetros de dos o varias distribuciones, sobre la independencia de las muestras, etc.

La hipótesis sobre el tipo de la distribución desconocida, o sobre los parámetros de las distribuciones conocidas se llama *estadística*.

Por ejemplo, serán estadísticas las siguientes hipótesis:

1) el conjunto general está distribuido según la ley de Poisson;

2) las dispersiones de dos conjuntos normales son iguales entre sí.

En la primera hipótesis se supuso sobre el tipo de la distribución desconocida, en la segunda, sobre los parámetros de dos distribuciones conocidas.

La hipótesis «en el año 1980 no habrá guerra» no es estadística, puesto que en ella no se trata ni del tipo ni de los parámetros de la distribución.

A la par de la hipótesis presentada se examina la hipótesis que la contradice. Si la hipótesis propuesta es rechazada, tiene lugar la hipótesis contradictoria. Por esta causa, conviene distinguir estas hipótesis.

Se llama nula (*fundamental*) la hipótesis H_0 presentada.

Se llama *concurrente* (*alternativa*) la hipótesis H_1 que contradice la hipótesis nula o fundamental.

Por ejemplo, si la hipótesis fundamental consiste en suponer que la esperanza matemática a de una distribución normal (gaussiana) es igual a 10, la hipótesis concurrente, en particular, puede consistir en la suposición de que $a \neq 10$. Resumidamente esto se escribe así:

$$H_0\colon a = 10;\ \ H_1\colon a \neq 10.$$

Se distinguen las hipótesis que contienen sólo una suposición y las de más de una.

La hipótesis se llama *simple* cuando contiene sólo una suposición. Por ejemplo, si λ es el parámetro de una distribución exponencial, la hipótesis $H_0\colon \lambda = 5$ es simple. La hipótesis H_0: la esperanza matemática de una distribución normal, igual a 3 (σ es conocida), es simple.

La hipótesis se llama *compleja* cuando está compuesta de un número finito o infinito de hipótesis simples. Por ejemplo, la hipótesis compleja $H\colon \lambda > 5$ está compuesta de una multitud de hipótesis simples de tipo $H_i : \lambda = b_i$, donde b_i es un número cualquiera mayor que 5. La hipótesis H_0: la esperanza matemática de una distribución normal, igual a 3 (σ incógnita), es compleja.

§ 2. Errores de primer y de segundo género

La hipótesis presentada puede ser cierta o errónea, por eso surge la necesidad de su verificación. Puesto que la verificación se realiza por métodos estadísticos, la misma se llama *estadística*. Debido a la verificación estadística de la hipótesis en dos casos se puede haber tomado una resolución incorrecta, es decir, pueden permitirse errores de dos géneros.

El *error de primer género* consiste en que será rechazada la hipótesis verdadera.

El *error de segundo género* consiste en que será admitida la hipótesis errónea.

Cabe hacer notar que las consecuencias de estos errores pueden ser muy diferentes. Por ejemplo, si se rechaza la solución correcta de «continuar la construcción de una vivienda», este error de primer género entraña una pérdida material; si se toma la resolución incorrecta de «continuar la construcción», a pesar del peligro de derrumbe de la construcción, este error de segundo género puede ocasionar la muerte de personas. Desde luego, podemos exponer ejemplos, en los cuales el error de primer género da lugar a consecuencias más graves que el error de segundo género.

Nota 1. También en dos casos se puede tomar una resolución correcta:

1) la hipótesis se admite; además, en realidad ella es correcta;
2) la hipótesis se rechaza; además, en realidad ella es falsa.

Nota 2. La probabilidad de cometer un error de primer género se designa por α, y se llama *nivel de significación*, el que con mayor frecuencia toma el valor 0,05 o bien 0,01. Si, por ejemplo, se toma el nivel de significación igual a 0,05, significa que en cinco casos de cien arriesgamos cometer un error de primer género (rechazar una hipótesis correcta).

§ 3. Criterio estadístico de verificación de la hipótesis nula. Valor observado del criterio

Para verificar la hipótesis nula (cero) o fundamental se utiliza una magnitud aleatoria especialmente escogida, cuya distribución exacta o aproximada es conocida. Esta magnitud se designa por U o bien Z, si está distribuida normalmente; por F o v^2, según la ley de Fisher—Snedecor; T, según la ley t de Student, χ^2, por la ley de «ji cuadrado», etc. Puesto que en este párrafo no se tomará en cuenta el

tipo de distribución, designamos esta magnitud por K a fin de generalizar.

Se llama *criterio estadístico* (o simplemente *criterio*) la magnitud aleatoria K que sirve para verificar la hipótesis fundamental.

Por ejemplo, si se verifica la hipótesis de igualdad de las dispersiones de dos conjuntos generales normales, como criterio K se toma la relación de las dispersiones muestrales corregidas:

$$F = \frac{s_1^2}{s_2^2}.$$

Esta magnitud es aleatoria, ya que en distintas pruebas las dispersiones tomarán distintos valores previamente desconocidos, y está distribuida por la ley de Fisher—Snedecor.

Para verificar la hipótesis por los datos de las muestras se calculan los valores particulares de las magnitudes que entran en el criterio, y, por consiguiente, se obtiene un valor particular (observado) del criterio.

Por *el valor observado* empírico K_{obs} se designa el valor del criterio, calculado por las muestras.

Por ejemplo, si por dos muestras extraídas de conjuntos generales normales, se hallan las dispersiones muestrales corregidas $s_1^2 = 20$ y $s_2^2 = 5$, el valor observado del criterio F es

$$F_{obs} = \frac{s_1^2}{s_2^2} = \frac{20}{5} = 4,$$

§ 4. Región crítica. Región de aceptación de la hipótesis. Puntos críticos

Después de elegir un criterio determinado, el conjunto de todos sus valores posibles se dividen en dos subconjuntos no intersecables: uno de ellos contiene valores del criterio, para los cuales la hipótesis nula se rechaza, y el otro, para los cuales ella se acepta.

El conjunto de valores del criterio, para los cuales la hipótesis fundamental se rechaza, se llama *región crítica*.

El conjunto de valores del criterio, para los cuales la hipótesis se acepta se llama *región de aceptación de la hipótesis* (región de valores admisibles).

El principio fundamental de verificación de la hipótesis estadística puede formularse así: si el valor observado del

criterio corresponde a la región crítica, la hipótesis se rechaza; si el valor observado del criterio corresponde a la región de aceptación de la hipótesis, la hipótesis se acepta.

Puesto que el criterio K es al mismo tiempo una magnitud aleatoria, todos sus valores posibles pertenecen a un cierto

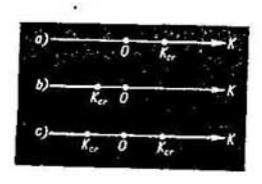


Fig. 23.

intervalo. Por eso, la región crítica y la región de aceptación de la hipótesis también son intervalos, y, por lo tanto, existen puntos que los separan.

Los puntos que separan la región crítica de la región de aceptación de la hipótesis se llaman *puntos críticos* (límites) k_{cr}.

Se distinguen las regiones críticas unilateral (de derecha o de izquierda) y bilateral

La región crítica determinada por la desigualdad $K > k_{cr}$, donde k_{cr} es un número positivo, se llama *de derecha* (fig. 23, *a*).

La región crítica determinada por la desigualdad $K < k_{cr}$, donde k_{cr} es un número negativo, se llama *de izquierda* (fig. 23, *b*).

La región crítica de derecha o de izquierda se llama *unilateral.*

La región crítica determinada por las desigualdades $K < k_1$, $K > k_2$, donde $k_2 > k_1$, se llama *bilateral.*

En particular, si los puntos críticos son simétricos respecto a cero, la región crítica bilateral se determina por las desigualdades (suponiendo que $k_{cr} > 0$):

$$K < -k_{cr}, \quad K > k_{cr},$$

o bien por la desigualdad equivalente $|K| > k_{cr}$ (fig. 23, *c*).

§ 5. Hallazgo de la región crítica de derecha

¿Cómo hallar una región crítica? Para responder argumentadamente a esta pregunta se requiere la aplicación de una teoría bastante compleja. Nos limitaremos a sus elementos. Para precisar, comencemos por hallar la región crítica de derecha que se determina por la desigualdad.

$$K > k_{cr},$$

donde $k_{cr} > 0$.

Vemos que para hallar la región crítica de derecha es suficiente encontrar el punto crítico. Por lo tanto, surge un nuevo problema: ¿cómo hallarlo?

Con este propósito se da una probabilidad bastante pequeña, o sea, el nivel de significación α. A continuación se busca el punto crítico k_{cr}, partiendo del requisito de que, si es válida la hipótesis fundamental, la probabilidad de que el criterio K tome un valor mayor que k_{cr} sea igual al nivel de significación aceptada:

$$P(K > k_{cr}) = \alpha.$$

Para cada criterio existen las tablas correspondientes por las cuales se halla el punto crítico que satisfaga este requisito.

Nota 1. Cuando ya se ha hallado el punto crítico, se calcula el valor observado del criterio por los datos de las muestras, y si resulta que $K_{obs} > k_{cr}$, se rechaza la hipótesis fundamental; si $K_{obs} < k_{cr}$, no hay porque rechazar la hipótesis fundamental (cero).

Explicación. ¿Por qué la región crítica de derecha se determinó, partiendo de la condición de que si la hipótesis fundamental es cierta, se cumple la correlación

$$P(K > k_{cr}) = \alpha? \qquad (*)$$

Puesto que la probabilidad del suceso $K > k_{cr}$ es pequeña (α es una probabilidad pequeña), al ser cierta la hipótesis fundamental, en virtud del principio de imposibilidad práctica de los sucesos poco probables, tal suceso no debe ocurrir en una sola prueba (cap. II, § 4). Si con todo éste ocurre, es decir, el valor observado del criterio resulta mayor que k_{cr}, ello es debido a que la hipótesis fundamental es falsa, y, por lo tanto, debe ser rechazada. De este modo, la condición (*) determina tales valores del criterio, para los cuales la hipótesis fundamental se rechaza, y que son los que componen, precisamente, la región crítica de derecha.

Nota 2. El valor observado del criterio puede resultar mayor que k_{cr} no porque la hipótesis fundamental es falsa, sino por otros motivos (pequeño volumen de la muestra, insuficiencia del método del experimento, etc.). En este caso, rechazando la hipótesis fundamental correcta, se comete un error de primer género. La probabilidad de este error es igual al nivel de significación α. Así, utilizando el requisito (*) con probabilidad α arriesgamos cometer un error de primer género.

Cabe hacer notar, a propósito, que en las obras respecto de control de la calidad de producción, la probabilidad de reconocer como defectuosa una partida de artículos buenos se llama «riesgo del productor», y la probabilidad de aceptar una partida defectuosa, «riesgo del consumidor».

Nota 3. Supongamos que se admite la hipótesis nula; es erróneo pensar que con ello queda demostrada. En efecto, se sabe que un ejemplo que confirme la validez de cierta afirmación general, aún no es su demostración. Por eso, es más correcto decir «los datos de las observaciones concuerdan con la hipótesis nula, y, por lo tanto, no dan motivos para rechazarla».

En la práctica, para aceptar una hipótesis con mayor certeza, ésta se verifica por otros métodos, o bien se repite el experimento aumentando el volumen de la muestra.

Una hipótesis se rechaza más categóricamente que se acepta. En realidad, se sabe que es suficiente dar un ejemplo que contradiga cierta afirmación general, para que esta afirmación se rechace. Si resultase que el valor observado del criterio corresponde a la región crítica, este hecho sirve precisamente de ejemplo contradictorio a la hipótesis fundamental, lo que permite rechazarla.

§ 6. Hallazgo de las regiones críticas de izquierda y bilateral

La búsqueda de las regiones críticas de izquierda y bilateral se reduce (al igual que para la región crítica de derecha) a hallar los correspondientes puntos críticos.

La región crítica de izquierda se determina (§ 4) por las desigualdades $K < k_{cr}$ ($k_{cr} < 0$).

El punto crítico se halla partiendo de la condición de que si la hipótesis fundamental es cierta, la probabilidad de que el criterio tome un valor menor que k_{cr}, sea igual al nivel de significación aceptado:

$$P(K < k_{cr}) = \alpha.$$

La región crítica bilateral se determina (§ 4) por las desigualdades $K < k_1$, $K > k_2$.

Los puntos críticos se hallan partiendo de la condición de que si la hipótesis fundamental es cierta, la suma de las probabilidades de que el criterio toma un valor menor que k_1

o mayor que k_2, sea igual al nivel de significación admitido:

$$P(K < k_1) + P(K > k_2) = \alpha. \qquad (*)$$

Está claro que los puntos críticos pueden ser elegidos por multitud de métodos. Si la distribución del criterio es simétrica con respecto a cero y existen fundamentos (por ejemplo, para aumentar la potencia *) para elegir los puntos simétricos con respecto a cero $-k_{cr}$ y k_{cr} ($k_{cr} > 0$) entonces

$$P(K < -k_{cr}) = P(K > k_{cr}).$$

Teniendo en cuenta la (*), obtenemos

$$P(k > k_{cr}) = \frac{\alpha}{2}.$$

Esta correlación sirve precisamente para hallar los puntos críticos de una región crítica bilateral.

Como ya se indicara (§ 5), los puntos críticos se hallan por las tablas correspondientes.

§ 7. Conocimientos suplementarios sobre la elección de la región crítica. Potencia del criterio

Hemos construido la región crítica partiendo de la exigencia de que, la probabilidad de que el criterio caiga en ella sea igual a α, a condición de que la hipótesis fundamental es cierta. Resulta conveniente también examinar la probabilidad de que el criterio caiga en la región crítica a condición de que la hipótesis fundamental es falsa, y, por lo tanto, cierta la concurrente.

La probabilidad de que el criterio caiga en la región crítica, a condición de que la hipótesis alternativa es cierta, se llama *potencia del criterio*. En otras palabras, la potencia del criterio es la probabilidad de que la hipótesis fundamental sea rechazada si es cierta la hipótesis concurrente.

Supongamos que para verificar la hipótesis se toma un nivel de significación determinado y la muestra tiene un volumen fijado. Queda al arbitrio la elección de la región crítica. Mostraremos que conviene construirla de manera que la potencia del criterio sea máxima.

Nos cercioramos previamente de que si la probabilidad del error de segundo género (aceptar una hipótesis errónea)

* La definición de la potencia está dado en el § 7.

es igual a β, la potencia es igual a $1-\beta$. En efecto, si β es la probabilidad de un error de segundo género, es decir, del suceso: «aceptación de la hipótesis fundamental, además de ser cierta la concurrente», la probabilidad del suceso opuesto: «rechazo de la hipótesis fundamental, además de ser cierta la hipótesis concurrente», es decir, la potencia del criterio es igual a $1-\beta$.

Supongamos que la potencia $1-\beta$ crece; por lo tanto, disminuye la probabilidad β de cometer un error de segundo género. De este modo, cuanto mayor es la potencia, tanto menor es la probabilidad de un error de segundo género.

Por consiguiente, si el nivel de significación ya se ha elegido, *la región crítica hay que construirla de manera que la potencia del criterio sea máxima.* El cumplimiento de esta condición asegura un error de segundo género mínimo, lo que indudablemente es deseable.

Nota 1. Puesto que la probabilidad del suceso «se ha cometido un error de segundo género» es igual a β, la probabilidad del suceso opuesto «no se ha cometido un error de segundo género» es igual a $1-\beta$, es decir, igual a la potencia del criterio. De aquí se deduce que la potencia del criterio es la probabilidad de que no se cometerá un error de segundo género.

Nota 2. Está claro que cuanto menor es la probabilidad de los errores de primer y segundo género, tanto «mejor» es la región crítica. Sin embargo, para un volumen dado de la muestra, *no es posible disminuir simultáneamente* α y β; si se reduce α, β crecerá. Por ejemplo, si se admite $\alpha = 0$, se aceptarán todas las hipótesis, incluso las falsas, es decir, crece la probabilidad β del error de segundo género.

¿Cómo elegir más conveniente el α? La respuesta depende de la «gravedad de las consecuencias» de los errores para cada problema concreto. Por ejemplo, si un error de primer género da lugar a grandes pérdidas, y uno de segundo género, a pequeñas pérdidas, hay que tomar α lo más pequeño posible.

Si α ya se ha elegido, utilizando el teorema de J. Neyman y E. Pearson, expuesto en cursos más completos, se puede construir la región crítica, para la cual β será mínima, y, por lo tanto, la potencia del criterio será máxima.

Nota 3. El único método para disminuir *simultáneamente* las probabilidades de los errores de primer y segundo género consiste en aumentar el volumen de las muestras.

§ 8. Comparación de dos dispersiones de conjuntos generales normales

En la práctica el problema de comparar las dispersiones se presenta, si hay que confrontar la precisión de aparatos, instrumentos, los propios métodos de mediciones, etc.

Evidentemente, es preferible aquel aparato, instrumento y método que asegure la dispersión mínima de los resultados de las mediciones.

Supongamos que los conjuntos generales X e Y están distribuidos normalmente. Por las muestras independientes de volúmenes n_1 y n_2, escogidas de estos conjuntos, se han hallado las dispersiones muestrales corregidas s_X^2 y s_Y^2. Por las dispersiones corregidas, para un nivel de significación dado α, hay que verificar la hipótesis fundamental, consistente en que las dispersiones generales de los conjuntos examinados son iguales entre sí:

$$H_0: D(X) = D(Y).$$

Teniendo en cuenta que las dispersiones corregidas son estimaciones no desviadas de las dispersiones generales (cap. XVI, § 13), es decir,

$$M[s_X^2] = D(X), \quad M[s_Y^2] = D(Y),$$

la hipótesis fundamental se puede escribir así:

$$H_0: M[s_X^2] = M[s_Y^2].$$

Por consiguiente, hay que verificar que las esperanzas matemáticas de las dispersiones muestrales corregidas son iguales entre sí. Este problema se plantea por que generalmente las dispersiones corregidas resultan distintas. Surge la pregunta: *¿las dispersiones corregidas se diferencian considerablemente o poco?*

Si resulta que la hipótesis fundamental (nula) es cierta, es decir, las dispersiones generales son iguales, la diferencia de las dispersiones corregidas es insignificante y se debe a causas fortuitas, en particular, a la selección aleatoria de los objetos de la muestra. Por ejemplo, si la diferencia de las dispersiones muestrales corregidas de los resultados de las mediciones, realizadas con dos aparatos, resulta insignificante, los aparatos tienen igual precisión.

Si la hipótesis nula se rechaza, es decir, las dispersiones generales son desiguales, la diferencia de las dispersiones corregidas es considerable y no puede ser debida a causas fortuitas, sino se debe a que las mismas dispersiones generales son distintas. Por ejemplo si la diferencia de las dispersiones muestrales corregidas de los resultados de las mediciones, realizadas con dos aparatos, es considerable, la precisión de los aparatos es distinta.

Como criterio de verificación de la hipótesis nula sobre la igualdad de las dispersiones generales, tomamos la relación entre la dispersión corregida mayor y la menor, es decir, la magnitud alcatoria.

$$F = \frac{s^2_{may}}{s^2_{men}}.$$

A condición de que la hipótesis fundamental es cierta, la magnitud F tiene la distribución de Fisher—Snedecor (cap. XII, § 15) con grados de libertad $k_1 = n_1 - 1$ y $k_2 = n_2 - 1$, donde n_1 es el volumen de la muestra, por el cual se ha calculado la mayor dispersión corregida, n_2 es el volumen de la muestra por la que se halló la menor dispersión.

Recordemos que la distribución de Fisher—Snedecor depende solamente del número de grados de libertad y no depende de otros parámetros.

La región crítica se construye en función del tipo de hipótesis alternativa.

PRIMER CASO. La hipótesis fundamental es $H_0: D(X) = D(Y)$. La hipótesis concurrente es $H_1: D(X) > D(Y)$.

En este caso se construye una región crítica de un solo lado,[1] precisamente de derecha, partiendo de la condición de que la probabilidad de que el criterio F caiga en esta región, suponiendo cierta la hipótesis fundamental, sea igual al nivel de significación aceptada:

$$P[F > F_{cr}(\alpha, k_1, k_2)] = \alpha.$$

El punto crítico $F_{cr}(\alpha, k_1, k_2)$ se halla por la tabla de los puntos críticos de la distribución de Fisher-Snedecor (suplemento 7), y entonces la región crítica de derecha se determina por la desigualdad

$$F > F_{cr},$$

y la región de aceptación de la hipótesis fundamental, por la desigualdad

$$F < F_{cr}.$$

La relación entre la dispersión corregida mayor y la menor, calculada por los datos de las observaciones, la designamos por F_{obs} y formulamos la regla de verificación de la hipótesis fundamental.

Regla 1. Para verificar, a un nivel de significación dado, la hipótesis fundamental $H_0: D(X) = D(Y)$ sobre la

igualdad de las dispersiones generales de conjuntos normales cuando la hipótesis concurrente es $H_1 : D(X) > D(Y)$, hay que calcular la relación entre las dispersiones corregidas mayor y menor, es decir,

$$F_{obs} = \frac{s^2_{may}}{s^2_{men}}$$

y por la tabla de puntos críticos de la distribución de Fisher—Snedecor, según el nivel de significación prefijado α y los números de grados de libertad k_1 y k_2 (k_1 es el número de grados de libertad de la dispersión corregida mayor), hallar el punto crítico $F_{obs}(\alpha, k_1, k_2)$.

Si $F_{obs} < F_{cr}$, no hay porque rechazar la hipótesis nula.

Si $F_{obs} > F_{cr}$, se rechaza la hipótesis nula.

Ejemplo 1. Por dos muestras independientes de volúmenes $n_1 = 12$ y $n_2 = 15$, extraídas de los conjuntos generales normales X e Y, se han hallado las dispersiones muestrales corregidas $s^2_X = 11{,}41$ y $s^2_Y = 6{,}52$. Para un nivel de significación de 0,05 verificar la hipótesis fundamental $H_0 : D(X) = D(Y)$ sobre la igualdad de las dispersiones generales, cuando la hipótesis concurrente $H_1 : D(X) > D(Y)$.

SOLUCION. Hallamos la relación entre las dispersiones corregidas mayor y menor:

$$F_{obs} = \frac{11{,}41}{6{,}52} = 1{,}75.$$

Puesto que la hipótesis concurrente tiene la forma $D(X) > D(Y)$, la región crítica es de derecha.

Por la tabla (suplemento 7), según el nivel de significación $\alpha = 0{,}05$ y los números de grados de libertad $k_1 = 12-1 = 11$ y $k_2 = 15-1 = 14$, hallamos el punto crítico $F_{cr}(0{,}05;\ 11;\ 14) = 2{,}57$.

Dado que $F_{obs} < F_{cr}$, no hay porque rechazar la hipótesis fundamental sobre la igualdad de las dispersiones generales.

SEGUNDO CASO. La hipótesis fundamental $H_0 : D(X) = D(Y)$. La hipótesis alternativa $H_1 : D(X) \neq D(Y)$.

En este caso se construye una región crítica bilateral, partiendo de la condición de que la probabilidad de que el criterio caiga en esta región, suponiendo que la hipótesis fun-

damental es cierta, sea igual al nivel de significación admitido α.

¿Cómo elegir los límites de la región crítica? Resulta que la potencia máxima (la probabilidad de que el criterio caiga en la región crítica, siendo cierta la hipótesis alternativa)



Fig. 24.

se logra cuando la probabilidad de que el criterio caiga en cada uno de los intervalos de la región crítica es igual a $\frac{\alpha}{2}$.

Por consiguiente, si designamos por F_1 el límite izquierdo de la región crítica y por F_2, el derecho, tienen que producirse las correlaciones (fig. 24);

$$P(F < F_1) = \frac{\alpha}{2}, \quad P(F > F_2) = \frac{\alpha}{2}.$$

Como vemos, es suficiente encontrar los puntos críticos para hallar la propia región crítica

$$F < F_1, \; F > F_2,$$

así como la región de aceptación de la hipótesis fundamental:

$$F_1 < F < F_2.$$

¿Comó hallar prácticamente los puntos críticos?

El punto crítico derecho $F_2 = F_{cr}\left(\frac{\alpha}{2}, k_1, k_2\right)$ se halla directamente por la tabla de puntos críticos de la distribución de Fisher—Snedecor según el nivel de significación $\frac{\alpha}{2}$ y los grados de libertad k_1 y k_2.

Pero esta tabla no contiene datos de los puntos críticos izquerdos, por lo cual es imposible hallar F_1 directamente por la tabla.

Existe un método que permite superar esta dificultad. Empero no lo describiremos, ya que el punto crítico izquierdo se puede dejar de buscar. Nos limitamos a exponer el modo de garantizar la caída del criterio F en la región

crítica bilateral con la probabilidad igual al nivel de significación admitido α.

Resulta suficiente hallar el punto crítico derecho F_2 para un nivel de significación dos veces menor que el dado. En tal caso no sólo la probabilidad de que el criterio caiga en la «parte derecha» de la región crítica (es decir, más a la derecha de F_2) es igual a $\frac{\alpha}{2}$, sino que la probabilidad de que este criterio caiga en la «parte izquierda» de la región crítica (es decir, más a la izquierda de F_1) será también igual a $\frac{\alpha}{2}$. Puesto que estos sucesos son mutuamente excluyentes, la probabilidad de que el criterio examinado caiga en toda la región crítica bilateral, será igual a $\frac{\alpha}{2}+\frac{\alpha}{2}=\alpha$.

Por consiguiente, cuando la hipótesis alternativa $H_1: D(X) \neq D(Y)$, es suficiente hallar el punto crítico $F_2 = F_{cr}\left(\frac{\alpha}{2}, k_1, k_2\right)$.

Regla 2. Para verificar, a un nivel de significación dado α, la hipótesis fundamental de la igualdad de las dispersiones generales de conjuntos distribuidos normalmente, cuando la hipótesis alternativa $H_1: D(X) \neq D(Y)$, hay que calcular la relación entre las dispersiones corregidas mayor y menor, es decir, $F_{obs} = \frac{s^2_{may}}{s^2_{men}}$ y por la tabla de puntos críticos de la distribución de Fisher—Snedecor según el nivel de significación $\frac{\alpha}{2}$ (dos veces menor que el dado) y por los números de grados de libertad k_1 y k_2 (k_1 es el número de grados de libertad de la dispersión mayor) hallar el punto crítico $F_{cr}\left(\frac{\alpha}{2}, k_1, k_2\right)$.

Si $F_{obs} < F_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $F_{obs} > F_{cr}$, la hipótesis fundamental se rechaza.

Ejemplo 2. Por dos muestras independientes de volúmenes $n_1 = 10$ y $n_2 = 18$, extraídas de los conjuntos generales normales X e Y, se han hallado las dispersiones muestrales corregidas $s^2_X = 1{,}23$ y $s^2_Y = 0{,}41$. Para el nivel de significación $\alpha = 0{,}1$, verificar la hipótesis fundamental de la igualdad de las dispersiones generales, cuando la hipótesis alternativa $H_1: D(X) \neq D(Y)$.

SOLUCION. Hallamos la relación entre las dispersiones corregidas mayor y menor:

$$F_{obs} = \frac{1,23}{0,41} = 3.$$

Por los datos la hipótesis alternativa (concurrente) tiene la forma $D(X) \neq D(Y)$, por eso la región crítica es bilateral.

Por la tabla, según el nivel de significación, dos veces menor que el prefijado, es decir, para $\frac{\alpha}{2} = \frac{0,1}{2} = 0,05$ y los números de grados de libertad $k_1 = 10-1 = 9$, $k_2 = 18-1 = 17$, hallamos el punto crítico $F_{cr}(0,05; 9; 17) = 2,50$.

Puesto que $F_{obs} > F_{cr}$, la hipótesis fundamental de la igualdad de las dispersiones generales la rechazamos. En otras palabras, las dispersiones muestrales corregidas se diferencian considerablemente. Por ejemplo, si las dispersiones examinadas caracterizaran la precisión de dos métodos de mediciones, hay que preferir el método que tiene una dispersión menor (0,41).

§ 9. Comparación de la dispersión muestral corregida con la dispersión general hipotética de un conjunto normal

Supongamos que un conjunto general está distribuido normalmente, pero existen motivos para suponer que la dispersión general, a pesar de ser desconocida, es igual al valor hipotético (supuesto) σ_0^2. En la práctica σ_0^2 se determina basándose en la experiencia anterior, o bien teóricamente.

Supongamos que de un conjunto general se ha extraído una muestra de volumen n y por ella se obtuvo la dispersión muestral corregida S^2 con $k = n - 1$ grados de libertad. Para un nivel de significación prefijado se necesita verificar por la dispersión corregida la hipótesis fundamental consistente en que la dispersión general del conjunto examinado es igual al valor hipotético σ_0^2.

Teniendo en cuenta que S^2 es la estimación no desviada de de la dispersión general, la hipótesis fundamental se puede escribir así:

$$H_0\colon M(S^2) = \sigma_0^2.$$

Así, se necesita verificar que la esperanza matemática de la dispersión corregida es igual al valor hipotético de la

dispersión general. En otras palabras, *hay que establecer si las dispersiones muestral corregida e hipotética generales se diferencian considerablemente o no.*

En la práctica, la hipótesis a examinar se verifica si hay que controlar la precisión de los aparatos, instrumentos, máquinas, métodos de investigación y estabilidad de procesos tecnológicos. Por ejemplo, si se conoce la característica admisible de dispersión de la dimensión a controlar de las piezas producidas por una máquina automática, igual a σ_0^2, y la dispersión corregida hallada por la muestra resulta de manera significativa mayor que σ_0^2, entonces hay que reajustar la máquina.

Como criterio de verificación de la hipótesis nula tomamos la magnitud alleartoria $\frac{(n-1)S^2}{\sigma_0^2}$. Esta magnitud es aleatoria, puesto que en distintos experimentos S^2 tomará distintos valores previamente desconocidos. Ya que se puede demostrar que tiene la distribución χ^2 con $k = n - 1$ grados de libertad (cap. XII. § 13). lo designamos por χ^2.

De este modo, el criterio de verificación de la hipótesis nula es

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}.$$

La región crítica se construye según el tipo de hipótesis alternativa (concurrente).

PRIMER CASO. La hipótesis fundamental $H_0: \sigma^2 = \sigma_0^2$. La hipótesis alternativa $H_1: \sigma^2 > \sigma_0^2$.

En este caso, la región crítica de derecha se construye partiendo de la condición de que la probabilidad de que el criterio caiga en esta región, suponiendo cierta la hipótesis fundamental, sea igual al nivel de significación admitido:

$$P[\chi^2 > \chi^2_{cr}(\alpha, k)] = \alpha.$$

El punto crítico $\chi^2_{cr}(\alpha, k)$ se halla por la tabla de puntos críticos de la distribución χ^2 (suplemento 5) y, luego la región crítica de derecha se determina por la desigualdad

$$\chi^2 > \chi^2_{cr},$$

mientras que la región de aceptación de la hipótesis fundamental, por la desigualdad

$$\chi^2 < \chi^2_{cr}.$$

Designemos el valor del criterio, calculado según los datos de las observaciones, por χ^2_{obs} y formulemos la regla de verificación de la hipótesis nula.

Regla 1. Para verificar, a un nivel de significación dado α, la hipótesis nula $H_0 : \sigma^2 = \sigma_0^2$ sobre la igualdad entre la dispersión general desconocida de un conjunto normal y el valor hipotético, cuando la hipótesis alternativa $H_1 : \sigma^2 > \sigma_0^2$, hay que calcular el valor observado del criterio $\chi^2_{obs} = \frac{(n-1)S^2}{\sigma_0^2}$ y por la tabla de puntos críticos de la distribución χ^2, según el nivel de significación prefijado α y el número de grados de libertad $k = n - 1$, hallar el punto crítico $\chi^2_{cr}(\alpha, k)$.

Si $\chi^2_{obs} < \chi^2_{cr}$ no hay porque rechazar la hipótesis nula.

Si $\chi^2_{obs} > \chi^2_{cr}$ la hipótesis nula se rechaza.

Ejemplo 1. De un conjunto general normal se ha extraído una muestra de volumen $n = 13$ y por ella se ha hallado la dispersión muestral corregida $s^2 = 14,6$. Se necesita verificar la hipótesis fundamental $H_0 : \sigma^2 = \sigma_0^2 = 12$, para el nivel de significación 0,01, tomando como hipótesis alternativa $H_1 : \sigma^2 > 12$.

SOLUCION. Hallamos el valor observado del criterio.

$$\chi^2_{obs} = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(13-1)\cdot 14,6}{12} = 14,6.$$

Por los datos la hipótesis alternativa tiene la forma $\sigma^2 > 12$, por lo tanto la región crítica es de derecha.

Por la tabla (suplemento 5), según el nivel de significación 0,01 y el número de grados de libertad $k = n - 1 = 13 - 1 = 12$, hallamos el punto crítico $\chi^2_{cr}(0,01; 12) = 26,2$.

Puesto que $\chi^2_{obs} < \chi^2_{cr}$, no hay motivos para rechazar la hipótesis fundamental. En otras palabras, la diferencia entre la dispersión corregida (14,6) y la dispersión general hipotética (12) no es significante.

SEGUNDO CASO. La hipótesis fundamental $H_0 : \sigma^2 = \sigma_0^2$. La hipótesis alternativa $H_1 : \sigma^2 \neq \sigma_0^2$.

En este caso, se construye la región crítica bilateral, partiendo de la condición de que la probabilidad de que el criterio caiga en esta región, suponiendo cierta la hipótesis nula, sea igual al nivel de significación admitido α.

Los puntos críticos, o sea, los límites izquierdo y derecho de la región crítica, se hallan a condición de que la probabilidad de que el criterio caiga en cada uno de los dos intervalos de la región crítica, sea igual a $\frac{\alpha}{2}$:

$$P\left[\chi^2 < \chi^2_{\text{cr. izq}}\left(\frac{\alpha}{2}, k\right)\right] = \frac{\alpha}{2},$$

$$P\left[\chi^2 > \chi^2_{\text{cr. der}}\left(\frac{\alpha}{2}, k\right)\right] = \frac{\alpha}{2}.$$

En la tabla de puntos críticos de la distribución χ^2 se indican solamente los puntos críticos «derechos», por eso surge una dificultad aparente en buscar los puntos críticos «izquierdos». Esta dificultad se salva fácilmente si se tiene en cuenta que los sucesos

$$\chi^2 < \chi^2_{\text{cr. izq}} \quad \text{y} \quad \chi^2 > \chi^2_{\text{cr. izq}}$$

son opuestos y, por lo tanto, la suma de sus probabilidades es igual a la unidad:

$$P\left(\chi^2 < \chi^2_{\text{cr. izq}}\right) + P\left(\chi^2 > \chi^2_{\text{cr. izq}}\right) = 1.$$

De aquí

$$P\left(\chi^2 > \chi^2_{\text{cr. izq}}\right) = 1 - P\left(\chi^2 < \chi^2_{\text{cr. izq}}\right) = 1 - \frac{\alpha}{2}.$$

Como vemos, el punto crítico izquierdo se puede buscar, como de derecho (lo que significa que se puede hallar por la tabla), partiendo de la condición de que la probabilidad de que el criterio caiga en el intervalo situado a la derecha de este punto, sea igual a $1 - \frac{\alpha}{2}$.

Regla 2. Para verificar, a un nivel de significación dado α la hipótesis nula de la igualdad entre dispersión general desconocida σ^2 de un conjunto normal y el valor hipotético σ_0^2, cuando la hipótesis alternativa $H_1 : \sigma^2 \neq \sigma_0^2$, hay que calcular el valor observado del criterio $\chi^2_{\text{obs}} = \frac{(n-1)S^2}{\sigma_0^2}$ y por la tabla hallar el punto crítico izquierdo $\chi^2_{\text{cr}}\left(1 - \frac{\alpha}{2}, k\right)$ y el derecho, $\chi^2_{\text{cr}}\left(\frac{\alpha}{2}, k\right)$.

Si $\chi^2_{\text{cr.izq}} < \chi^2_{\text{obs}} < \chi^2_{\text{cr.der}}$, no hay porque rechazar la hipótesis fundamental.

Si $\chi^2_{\text{obs}} < \chi^2_{\text{cr.izq}}$ o bien $\chi^2_{\text{obs}} > \chi^2_{\text{cr.der}}$, la hipótesis fundamental se rechaza.

Ejemplo 2. De un conjunto general normal se ha extraído una muestra de volumen $n = 13$ y por ella se ha hallado la dispersión muestral corregida $s^2 = 10{,}3$. Se necesita verificar, para el nivel de significación 0,02, la hipótesis fundamental H_0: $\sigma^2 = \sigma_0^2 = 12$, tomando como hipótesis alternativa H_1: $\sigma^2 \neq 12$.

SOLUCION. Hallamos el valor observado del criterio

$$\chi^2_{obs} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(13-1)\cdot 10{,}3}{12} = 10{,}3.$$

Puesto que la hipótesis alternativa tiene la forma $\sigma^2 \neq 12$, la región crítica es bilateral.

Por la tabla (suplemento 5) hallamos los puntos críticos: izquierdo — $\chi^2_{cr}\left(1 - \frac{\alpha}{2},\ k\right) = \chi^2_{cr}\left(1 - \frac{0{,}02}{2},\ 12\right) = \chi^2_{cr}(0{,}99;\ 12) = 3{,}57$ y derecho — $\chi^2_{cr}\left(\frac{\alpha}{2},\ k\right) = \chi^2_{cr}(0{,}01;\ 12) = 26{,}2$.

Ya que el valor observado del criterio corresponde a la región de aceptación de la hipótesis $(3{,}57 < 10{,}3 < 26{,}2)$, no hay porque rechazarla. En otras palabras, la dispersión muestral corregida (10,3) se diferencia poco de la dispersión general hipotética (12).

CASO 3. La hipótesis alternativa es H_1: $\sigma^2 < \sigma_0^2$.

Regla 3. Para la hipótesis alternativa H_1: $\sigma^2 < \sigma_0^2$, se halla el punto crítico $\chi^2_{cr}(1 - \alpha,\ k)$.

Si $\chi^2_{obs} > \chi^2_{cr}(1 - \alpha,\ k)$, no hay porque rechazar la hipótesis fundamental.

Si $\chi^2_{obs} < \chi^2_{cr}(1 - \alpha,\ k)$ la hipótesis fundamental se rechaza.

Nota 1. Si se ha hallado la dispersión muestral D_m, como criterio se toma la magnitud aleatoria

$\chi^{2\prime} = \frac{nD_m}{\sigma_0^2}$ que tiene distribución χ^2 con $k = n - 1$ grados de libertad, bien se pasa a $s^2 = \frac{n}{n-1} D_m$.

Nota 2. Si el número de grados de libertad $k > 30$, el punto crítico $\chi^2_{cr}(\alpha,\ k)$ se puede hallar aproximadamente por la igualdad de Wilson — Hilferty

$$\chi^2_{cr}(\alpha,\ k) = k\left[1 - \frac{2}{9k} + Z_\alpha \cdot \sqrt{\frac{2}{9k}}\right]^3,$$

donde Z_α se halla, utilizando la función de Laplace (suplemento 2), por la igualdad $\Phi(Z_\alpha) = \frac{1-2\alpha}{2}$.

§ 10. Comparación de dos medias de conjuntos generales normales, cuyas dispersiones son conocidas (muestras independientes)

Supongamos que los conjuntos generales X e Y están distribuidos normalmente, asimismo se conocen sus dispersiones (por ejemplo, del experimento anterior, o bien hallados teóricamente). Por las muestras independientes de volúmenes n y m, extraídas de estos conjuntos, se han hallado las medias muestrales $\overline{x}$ e $\overline{y}$.

Por las medias muestrales y dado el nivel de significación α hay que verificar la hipótesis nula consistente en que las medias generales (esperanzas matemáticas) de los conjuntos a examinar son iguales entre sí, es decir,

$$H_0: \ M(X) = M(Y).$$

Teniendo en cuenta que las medias muestrales son estimaciones no desviadas de las medias generales (cap. XV, § 5), es decir, $M(\overline{X}) = M(X)$ y $M(\overline{Y}) = M(Y)$, la hipótesis nula puede escribirse así:

$$H_0: \ M(\overline{X}) = M(\overline{Y}).$$

Por consiguiente, es necesario verificar que las esperan as matemáticas de las medias muestrales son iguales entre sí. Este problema se plantea por que, como regla, las medias muestrales resultan distintas. Surge la pregunta: ¿las medias muestrales se diferencian de un modo significativo o insignificativo?

Si la hipótesis nula (cero) resulta cierta, es decir, las medias generales son iguales, la diferencia de las medias muestrales es insignificativa y se debe a causas fortuitas, en particular, a la selección aleatoria de los objetos de la muestra.

Por ejemplo, si las magnitudes físicas A y B tienen idénticas dimensiones verdaderas, y las medias aritméticas $\overline{x}$ e $\overline{y}$ de los resultados de las mediciones de estas magnitudes son distintas, esta diferencia es insignificativa.

Si la hipótesis nula se rechaza, es decir, las medias generales no son iguales, la diferencia de las medias muestrales es significativa y no puede ser debida a causas fortuitas, sino se debe a que las propias medias generales (esperanzas matemáticas) son diferentes. Por ejemplo, si

la media aritmética $\bar{x}$ de los resultados de las mediciones de la magnitud física A se diferencia de manera significativa de la media artitmética $\bar{y}$ de los resultados de las mediciones de la maguitud física B, esto demuestra que las dimensiones verdaderas (esperanzas matemáticas) de estas magnitudes son distintas.

Como criterio de verificación de la hipótesis nula tomamos la magnitud aleatoria

$$Z = \frac{\overline{X} - \overline{Y}}{\sigma(\overline{X} - \overline{Y})} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{D(X)}{n} + \frac{D(Y)}{m}}}.$$

Esta es una magnitud aleatoria, ya que en distintas pruebas $\bar{x}$ e $\bar{y}$ toman distintos valores previamente desconocidos.

EXPLICACION. Por definición de la desviación cuadrática media $\sigma(\overline{X}-\overline{Y}) = \sqrt{D(\overline{X}-\overline{Y})}$.

Basándonos en la propiedad 4 (cap. VIII, § 5) $D(\overline{X}-\overline{Y}) = D(\overline{X}) + D(\overline{Y})$.

Por la fórmula (*) (cap. VIII, § 9): $D(\overline{X}) = \frac{D(X)}{n}$, $D(\overline{Y}) = \frac{D(Y)}{n}$. Por lo tanto,

$$\sigma(\overline{X} - \overline{Y}) = \sqrt{\frac{D(X)}{n} + \frac{D(Y)}{m}}.$$

El criterio Z es una magnitud aleatoria normal normada. En efecto, la magnitud Z está distribuida normalmente, puesto que es la combinación lineal de las magnitudes $\overline{X}$ e $\overline{Y}$ normalmente distribuidas; estas propias magnitudes están distribuidas normalmente como las medias muestrales halladas por las muestras escogidas de los conjuntos generales normales; Z es una magnitud normada, porque $M(Z) = 0$, cuando la hipótesis fundamental es cierta, $\sigma(Z) = 1$, pues las muestras son independientes.

La región crítica se construye según el tipo de hipótesis alternativa.

PRIMER CASO. La hipótesis nula H_0: $M(X) = M(Y)$. La hipótesis alternativa H_1: $M(X) \neq M(Y)$.

En este caso se contruye la región crítica bilateral, partiendo de la condición de que la probabilidad de caer el

criterio en esta región, suponiendo que la hipótesis fundamental es cierta, sea igual al nivel de significación admitido α.

La potencia máxima del criterio (probabilidad de que el criterio caiga en la región crítica, cuando la hipótesis alternativa es cierta) se logra cuando los puntos críticos «izquierdo» y «derecho» se escogen de manera que la probabilidad de caer el criterio en cada uno de los dos intervalos de la región crítica, sea igual a $\frac{\alpha}{2}$:

$$P(Z < z_{cr.\,izq}) = \frac{\alpha}{2}, \qquad (*)$$
$$P(Z > z_{cr.\,der}) = \frac{\alpha}{2}.$$

Puesto que Z es una magnitud normal normada, y la distribución de tal magnitud es simétrica con respecto a cero, los puntos críticos también son simétricos con respecto a cero.



Fig. 25.

Por consiguiente, si el límite derecho de la región crítica bilateral la designamos por z_{cr}, tendremos que el límite izquierdo es igual a $-z_{cr}$ (fig. 25).

De este modo, es suficiente hallar el límite derecho para encontrar la propia región crítica bilateral

$$Z < -z_{cr}, \quad Z > z_{cr}$$

y la región de aceptación de la hipótesis nula es

$$(-z_{cr}, z_{cr}).$$

Veamos como hallar z_{cr}, o sea, el límite derecho de la región crítica bilateral, utilizando la función de Laplace $\Phi(z)$. Se sabe que la función de Laplace determina la probabilidad de que una magnitud aleatoria normal normada, por ejemplo Z, caiga en el intervalo $(0, z)$:

$$P(0 < Z < z) = \Phi(z). \qquad (**)$$

Dado que la distribución de Z es simétrica con respecto a cero, la probabilidad de que Z caiga en el intervalo $(0, \infty)$ es igual a $\frac{1}{2}$. Por lo tanto, si se descompone este intervalo mediante el punto z_{cr}, en los intervalos $(0, z_{cr})$ y (z_{cr}, ∞), por el teorema de la adición

$$P(0 < Z < z_{cr}) + P(Z > z_{cr}) = \frac{1}{2}. \qquad (***)$$

En virtud de (*) y (**), obtenemos

$$\Phi(z_{cr}) + \frac{\alpha}{2} = \frac{1}{2}.$$

Por lo tanto,

$$\Phi(z_{cr}) = \frac{1-\alpha}{2}.$$

De aquí deducimos: para hallar el límite derecho de la región crítica bilateral (z_{cr}) es suficiente encontrar el valor del argumento de la función de Laplace, al que corresponde un valor de la función, igual a $\frac{1-\alpha}{2}$.

En tal caso, la región crítica bilateral se determina por las desigualdades.

$$Z < -z_{cr}, \; Z > z_{cr}$$

o por la desigualdad equivalente

$$|Z| > z_{cr},$$

y la región de aceptación de la hipótesis nula por la desigualdad

$$-z_{cr} < Z < z_{cr},$$

o por la desigualdad equivalente

$$|Z| < z_{cr}.$$

Designemos el valor del criterio, calculado por los
de las observaciones, por Z_{obs} y formulemos la regla de veri-
datos ficación de la hipótesis nula.

Regla 1. Para verificar a un nivel de significación dado α la hipótesis nula H_0: $M(X) = M(Y)$ sobre la igualdad de las esperanzas matemáticas de dos conjuntos generales normales con dispersiones conocidas, cuando la hipótesis alternativa H_1: $M(X) \neq M(Y)$, hay que calcular

el valor observado del criterio $Z_{obs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{D(X)}{n} + \frac{D(Y)}{m}}}$ y por la tabla de la función de Laplace hallar el punto crítico por la igualdad $\Phi(z_{cr}) = \frac{1-\alpha}{2}$.

Si $|Z_{obs}| < z_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $|Z_{obs}| > z_{cr}$, la hipótesis fundamental se rechaza.

Ejemplo 1. Por dos muestras independientes de volúmenes $n = 60$ y $m = 50$, extraídas de conjuntos generales normales, se han hallado las medias muestrales $\bar{x} = 1250$ e $\bar{y} = 1275$. Las dispersiones generales son conocidas: $D(X) = 120$, $D(Y) = 100$. Para el nivel de significación 0,01, verificar la hipótesis fundamental H_0: $M(X) = M(Y)$, cuando la hipótesis alternativa H_1: $M(X) \neq M(Y)$.

SOLUCION. Hallamos el valor observado del criterio

$$Z_{obs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{D(X)}{n} + \frac{D(Y)}{m}}} = \frac{1250 - 1275}{\sqrt{\frac{120}{60} + \frac{100}{50}}} = -12,5.$$

Por los datos, la hipótesis alternativa tiene la forma $M(X) \neq M(Y)$, por eso la región crítica es bilateral.

Hallamos el punto crítico derecho por la igualdad

$$\Phi(z_{cr}) = \frac{1-\alpha}{2} = \frac{1-0,01}{2} = 0,495.$$

Por la tabla de la función de Laplace (suplemento 2) hallamos $z_{cr} = 2,58$.

Puesto que $|Z_{obs}| > z_{cr}$, rechazamos la hipótesis fundamental. En otras palabras, las medias muestrales se diferencian de modo significativo.

SEGUNDO CASO. La hipótesis fundamental H_0: $M(X) = M(Y)$. La hipótesis alternativa H_1: $M(X) > M(Y)$.

En la práctica este caso tiene lugar si las razones profesionales permiten suponer que la media general de un conjunto es mayor que la media general del otro. Por ejemplo, si se ha introducido el perfeccionamiento de un proceso tecnológico, es natural admitir que ello dará lugar al aumento de la producción.

En este caso se construye la región crítica de derecha partiendo de la condición de que la probabilidad de que el cri-



Fig. 26.

terio caiga en esta región suponiendo que la hipótesis fundamental es cierta, sea igual al nivel de significación admitido (fig. 26):

$$P(Z > z_{cr}) = \alpha. \qquad (****)$$

Mostremos cómo hallar el punto crítico mediante la función de Laplace. Utilizamos la correlación (***):

$$P(0 < Z < z_{cr}) + P(Z > z_{cr}) = \frac{1}{2}.$$

En virtud de (**) y (****) tenemos

$$\Phi(z_{cr}) + \alpha = \frac{1}{2}.$$

Por lo tanto,

$$\Phi(z_{cr}) = \frac{1-2\alpha}{2}.$$

De aquí deducimos que para hallar el límite de la región crítica de derecha (z_{cr}), es suficiente encontrar el valor del argumento de la función de Laplace, al que corresponde el valor de la función, igual a $\frac{1-2\alpha}{2}$. En tal caso, la región crítica de derecha se determina por la desigualdad $Z > z_{cr}$, y la región de aceptación de la hipótesis fundamental (nula), por la desigualdad $Z < z_{cr}$.

Regla 2. Para verificar a un nivel de significación prefijado α, la hipótesis fundmental H_0: $M(X) = M(Y)$ de la igualdad entre las esperanzas matemáticas de dos conjuntos generales normales de dispersiones conocidas, cuando la hipótesis alternativa H_1: $M(X) > M(Y)$, hay que calcular el valor observado del criterio $Z_{obs} = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{D(X)}{n}+\frac{D(Y)}{m}}}$

y por la tabla de función de Laplace hallar el punto crítico de la igualdad $\Phi(z_{cr}) = \frac{1-2\alpha}{2}$.

Si $Z_{obs} < z_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $Z_{obs} > z_{cr}$, la hipótesis fundamental se rechaza.

Ejemplo 2. Por dos muestras independientes de volúmenes $n = 10$ y $m = 10$, escogidas de conjuntos generales normales, se han hallado las medias muestrales $\bar{x} = 14,3$ e $\bar{y} = 12,2$. Las dispersiones generales son conocidas: $D(X) = 22$, $D(Y) = 18$. Verificar la hipótesis fundamental H_0: $M(X) = M(Y)$ para el nivel de significación 0,05, siendo la hipótesis alternativa H_1: $M(X) > M(Y)$.

SOLUCION. Hallamos el valor observado del criterio

$$Z_{obs} = \frac{14,3-12,2}{\sqrt{\frac{22}{10}+\frac{18}{10}}} = 1,05.$$

Por los datos, la hipótesis alternativa tiene la forma $M(X) > M(Y)$, por eso la región crítica es de derecha.

Por la tabla de la función de Laplace hallamos que $z_{cr} = 1,64$.

Puesto que $Z_{obs} < z_{cr}$, no hay porque rechazar la hipótesis fundamental. En otras palabras, las medias muestrales se diferencian de manera insignificativa.

TERCER CASO. La hipótesis fundamental H_0: $M(X) = M(Y)$. La hipótesis alternativa H_1: $M(X) < M(Y)$.

En este caso se construye la región crítica de izquierda, partiendo de la condición de que la probabilidad de caer el



Fig. 27.

criterio en esta región, suponiendo que la hipótesis fundamental es cierta, sea igual al nivel de significación admitido (fig. 27):

$$P(Z < z'_{cr}) = \alpha.$$

Tomando en consideración que el criterio Z está distribuido simétricamente respecto a cero, deducimos que el punto crítico buscado z'_{cr} es simétrico de un punto tal $z_{cr} > 0$, para el que $P(Z > z_{cr}) = \alpha$, es decir, $z'_{cr} = -z_{cr}$. Por consiguiente, para hallar el punto z_{cr}, es suficiente al principio hallar el «punto auxiliar» z'_{cr} del modo descripto en el *segundo caso*, y luego tomar el valor encontrado con signo menos. Entonces la región crítica de izquierda se determina por la desigualdad $Z < -z_{cr}$, y la región de aceptación de la hipótesis fundamental, por la desigualdad $Z > -z_{cr}$.

Regla 3. Para la hipótesis alternativa H_1: $M(X) < M(Y)$ hay que calcular Z_{obs} y, al comienzo, por la tabla de la función de Laplace hay que hallar el «punto auxiliar» z_{cr} mediante la igualdad $\Phi(z_{cr}) = \frac{1-2\alpha}{2}$, y luego poner $z'_{cr} = -z_{cr}$.

Si $Z_{obs} > -z_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $Z_{obs} < -z_{cr}$, la hipótesis fundamental se rechaza.

Ejemplo 3. Por dos muestras independientes de volúmenes $n = 50$ y $m = 50$, extraídas de conjuntos generales normales, se han hallado las medias muestrales $\overline{x} = 142$ e $\overline{y} = 150$. Las dispersiones generales son conocidas: $D(X) = 28{,}2$, $D(Y) = 22{,}8$. Verificar la hipótesis fundamental H_0: $M(X) = M(Y)$ para el nivel de significación 0,01, siendo la hipótesis alternativa H_1: $M(X) < M(Y)$.

SOLUCION. Sustituyendo los datos del problema en la fórmula para calcular el valor observado del criterio, obtenemos $Z_{obs} = -8$.

Según los datos, la hipótesis alternativa tiene la forma $M(X) < M(Y)$, por eso la región crítica es de izquierda Hallamos el «punto auxiliar» z_{cr} mediante la igualdad

$$\Phi(z_{cr}) = \frac{1-2\alpha}{2} = \frac{1-2\cdot 0{,}01}{2} = 0{,}49.$$

Por la tabla de la función de Laplace hallamos $z_{cr} = 2{,}33$. Por lo tanto, $z'_{cr} = -z_{cr} = -2{,}33$.

Como $Z_{obs} < -z_{cr}$, rechazamos la hipótesis fundamental. En otras palabras, la media muestral $\overline{x}$ es bastante menor que la media moestral $\overline{y}$.

§ 11. Comparación de dos medias de conjuntos generales arbitrariamente distribuidos (grandes muestras independientes)

En el párrafo precedente se supuso que los conjuntos generales X e Y están distribuidos normalmente, y sus dispersiones son conocidas. Con esta suposición, cuando la hipótesis fundamental de la igualdad entre las medias para muestras independientes, el criterio Z está distribuido con precisión normalmente con parámetros 0 y 1.

Si no se cumple aunque sea una de las condiciones expuestas, el método comparativo de las medias, descripto en el § 10, no es aplicable.

Sin embargo, si las muestras independientes tienen un gran volumen (no menos de 30 cada una), las medias muestrales están distribuidas con aproximación normalmente, mientras que las dispersiones muestrales son estimaciones bastante buenas de las dispersiones generales y en este sentido se pueden considerar conocidas con aproximación. En resumen el criterio

$$Z' = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{D_m(X)}{n}+\frac{D_m(Y)}{m}}}$$

está distribuido con aproximación normalmente con parámetros $M(Z')=0$ (a condición de que la hipótesis fundamental sea cierta) y $\sigma(Z')=1$ (si las muestras son independientes).

De este modo, si: 1) dos conjuntos generales están distribuidos normalmente, y sus dispersiones son desconocidas, 2) los conjuntos generales no están distribuidos normalmente, y sus dispersiones son conocidas; 3) los conjuntos generales no están distribuidos normalmente y sus dispersiones son desconocidas; además, las muestras tienen un gran volumen y son independientes, podemos comparar las medias como se ha descripto en el § 10, sustituyendo el criterio exacto Z por el criterio aproximado Z'. En este caso, el valor observado del criterio aproximado es:

$$Z'_{obs} = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{D_m(X)}{n}+\frac{D_m(Y)}{m}}}.$$

Nota. Puesto que el criterio examinado es aproximado, conviene tener cuidado con las deducciones obtenidas por este criterio.

Ejemplo. Por dos muestras independientes de volúmenes $n = 100$ y $m = 120$, se han obtenido las medias muestrales $\overline{x} = 32{,}4$, $\overline{y} = 30{,}1$ y las dispersiones muestrales $D_m(X) = 15{,}0$, $D_m(Y) = 25{,}2$. Hay que verificar la hipótesis fundamental H_0: $M(X) = M(Y)$ para el nivel de significación 0,05, siendo la hipótesis alternativa $H_1 : M(X) \neq M(Y)$.

SOLUCIÓN. Sustituyendo los datos del problema en la fórmula para el cálculo del valor observado del criterio aproximado, obtenemos $Z_{obs} = 3{,}83$.

Según los datos, la hipótesis alternativa tiene la forma $M(X) > M(Y)$, por eso la región crítica es de derecha.

Hallamos el punto crítico por la igualdad

$$\Phi(z_{cr}) = \frac{1-2\alpha}{2} = \frac{1-2\cdot 0{,}05}{2} = 0{,}45.$$

Por la tabla de la función de Laplace hallamos que $z_{cr} = 1{,}64$.

Ya que $Z_{obs} > z_{cr}$, rechazamos la hipótesis fundamental. En otras palabras, las medias muestrales se diferencian de manera significativa.

§ 12. Comparación de dos medias de conjuntos generales normales, cuyas dispersiones son desconocidas e idénticas (pequeñas muestras independientes)

Supongamos que los conjuntos generales X e Y están distribuidos normalmente y sus dispersiones son desconocidas. Por ejemplo, por las muestras de pequeño volumen no se pueden obtener buenas estimaciones de las dispersiones generales. Por este motivo, el método comparativo de las medias, expuesto en el § 11, no es aplicable.

Sin embargo, si suponemos complementariamente que *las dispersiones generales desconocidas son iguales entre sí*, se puede formar el criterio comparativo de medias (de Student). Por ejemplo, si se comparan las dimensiones medias de dos partidas de piezas producidas en una misma máquina herramienta, es natural admitir que las dispersiones de las dimensiones a controlar son idénticas.

Si no hay motivos para considerar que las dispersiones son idénticas, *antes de comparar las medias* debe verificarse previamente la hipótesis de la igualdad entre las dispersiones generales, utilizando el criterio de Fisher—Snedecor (§ 8).

De este modo, suponiendo que las dispersiones generales son idénticas, es necesario verificar la hipótesis fundamental $H_0 : M(X) = M(Y)$. En otras palabras, hay que establecer si las medias muestrales $\overline{x}$ e $\overline{y}$ halladas por pequeñas muestras independientes de volúmenes n y m, se diferencian de manera significativa o insignificativa.

Como criterio de verificación de la hipótesis fundamental tomamos la magnitud aleatoria

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{(n-1) S_X^2 + (m-1) S_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Se ha demostrado que la magnitud T, cuando la hipótesis fundamental es cierta, tiene la distribución t de Student con $k = n + m - 2$ grados de libertad.

La región crítica se construye en función del tipo de hipótesis alternativa.

PRIMER CASO. La hipótesis fundamental $H_0: M(X) = M(Y)$. La hipótesis alternativa $H_1: M(X) \neq M(Y)$.

En este caso se construye la región crítica bilateral, partiendo de la condición de que la probabilidad de caer el criterio T en esta región, suponiendo que la hipótesis fundamental es cierta, será igual al nivel de significación admitido α.

La potencia máxima del criterio (probabilidad de que el criterio caiga en la región crítica, si la hipótesis alternativa es cierta) se logra cuando los puntos críticos «izquierdo» y «derecho» se eligen de tal modo que la probabilidad de que el criterio caiga en cada uno de los dos intervalos de la región crítica bilateral, es igual a $\frac{\alpha}{2}$:

$$P(T < t_{\text{cr. izq}}) = \frac{\alpha}{2}, \quad P(T > t_{\text{cr. der}}) = \frac{\alpha}{2}.$$

En virtud de que la magnitud T tiene la distribución t de Student, y ella es simétrica respecto a cero, tendremos que también los puntos críticos son simétricos respecto a cero. Por con siguiente, si designamos el límite derecho de la región crítica bilateral por $t_{\text{cr.bilat}}(\alpha, k)$, el límite izquierdo será igual a $-t_{\text{cr.bilat}}(\alpha, k)$. Así pues, es suficiente hallar el límite derecho de la región crítica para encontrar la región crítica bilateral,

$$T < -t_{\text{cr.bilat}}(\alpha, k), \quad T > t_{\text{cr.bilat}}(\alpha, k)$$

y la región de aceptación de la hipótesis fundamental

$$[-t_{cr.bilat}(\alpha, k),\ t_{cr.bilat}(\alpha, t)].$$

Designemos el valor del criterio, calculado por los datos de las observaciones, por T_{obs} y formulemos la regla de verificación de la hipótesis fundamental.

Regla 1. Para verificar, a un nivel de significación prefijado α, la hipótesis fundamental H_0: $M(X) = M(Y)$ de la igualdad entre las esperanzas matemáticas de dos conjuntos normales con dispersiones desconocidas, pero idénticas (en el caso de pequeñas muestras independientes), cuando la hipótesis alternativa H_1: $M(X) \neq M(Y)$, hay que calcular el valor observado del criterio

$$T_{obs} = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}$$

y según la tabla de puntos críticos de la distribución t de Student, por el nivel de significación dado α (dispuesto en la línea superior de la tabla) y el número de grados de libertad $k = n + m - 2$, hallar el punto crítico $t_{cr.bilat}(\alpha, k)$.

Si $|T_{obs}| < t_{cr.bilat}(\alpha, k)$, no hay porque rechazar la hipótesis fundamental.

Si $|T_{obs}| > t_{cr.bilat}(\alpha, k)$, la hipótesis fundamental se rechaza.

Ejemplo. Por dos muestras pequeñas independientes de volúmenes $n = 5$ y $m = 6$, escogidas de los conjuntos generales normales X e Y, se han hallado las medias muestrales $\bar{x} = 3{,}3$, $\bar{y} = 2{,}48$ y las dispersiones corregidas $s_X^2 = 0{,}25$ y $s_Y^2 = 0{,}108$. Para el nivel de significación 0,05, verificar la hipótesis fundamental H_0: $M(X) = M(Y)$, siendo la hipótesis alternativa H_1: $M(X) \neq M(Y)$.

SOLUCIÓN. Puesto que las dispersiones muestrales son distintas, verificamos previamente la hipótesis fundamental de la igualdad entre las dispersiones generales, utilizando el criterio de Fisher—Snedecor (§ 8).

Hallamos la relación entre las dispersiones corregidas mayor y menor

$$F_{obs} = \frac{0{,}25}{0{,}108} = 2{,}31.$$

La dispersión s_X^2 es bastante mayor que la dispersion s_Y^2, por eso tomamos como hipótesis alternativa H_1: $D(X) > D(Y)$. En este caso, la región crítica es de derecha. Según

la tabla, por el nivel de significación $\alpha = 0{,}05$ y los números de grados de libertad $k_1 = 5 - 1 = 4$, $k_2 = 6 - 1 = 5$, hallamos el punto crítico F_{cr} (0,05; 4; 5) = 5,19.

Puesto que $F_{obs} < F_{cr}$, no hay porque rechazar la hipótesis fundamental de la igualdad entre las dispersiones generales.

Ya que la hipótesis de la igualdad de las dispersiones generales se cumple, comparamos las medias.

Calculamos el valor observado del criterio t de Student:

$$T_{obs} = \frac{\bar{x} - \bar{y}}{\sqrt{ns_X^2 + ms_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Poniendo los valores numéricos de las magnitudes que entran en esta fórmula, obtenemos $T_{obs} = 3{,}27$.

Por los datos, la hipótesis alternativa tiene la forma $M(X) \neq M(Y)$, por eso la región crítica es bilateral. Por el nivel de significación 0,05 y el número de grados de libertad $k = 5 + 6 - 2 = 9$, hallamos según la tabla (suplemento 6) el punto crítico $t_{cr.\,bilat}$ (0,05; 9) = 2,26.

Puesto que $T_{obs} > t_{cr.\,bilat}$, rechazamos la hipótesis fundamental de la igualdad entre las medias generales. En otras palabras, las medias muestrales se diferencian de manera significativa.

SEGUNDO CASO. La hipótesis fundamental H_0: $M(X) = M(Y)$. La hipótesis alternativa H_1: $M(X) > M(Y)$.

En este caso se construye la región crítica de derecha, partiendo de la condición de que la probabilidad de caer el criterio T en esta región suponiendo que la hipótesis fundamental es cierta, sea igual al nivel de significación admitido:

$$P(T > t_{cr.\,der}) = \alpha.$$

El punto crítico $t_{cr.\,der}(\alpha, k)$ se halla por la tabla (suplemento 6) según el nivel de significación α situado en la línea inferior de la tabla y por el número de grados de libertad $k = n + m - 2$.

Si $T_{obs} < t_{cr.\,der}$, no hay porque rechazar la hipótesis fundamental.

Si $T_{obs} > t_{cr.\,der}$, la hipótesis fundamental se rechaza.

TERCER CASO. La hipótesis fundamental H_0: $M(X) = M(Y)$. La hipótesis alternativa H_1: $M(X) < M(Y)$.

En este caso, se construye la región crítica de izquierda, partiendo de la condición de que la probabilidad de caer el

criterio en esta región, suponiendo que la hipótesis fundamental es cierta, sea igual al nivel de significación α admitido:

$$P(T < t_{cr\ izq}) = \alpha.$$

En virtud de la simetría de la distribución de Student respecto a cero

$$t_{cr.\ izq} = -t_{cr.\ der}.$$

Por eso, al principio hallamos el punto crítico «auxiliar» $t_{cr.\ der}$, como se describe en el *segundo caso* y admitimos que $t_{cr\ izq} = -t_{cr.\ der}$.

Si $T_{obs} > -t_{cr.\ der}$, no hay porque rechazar la hipótesis fundamental.

Si $T_{obs} < t_{cr\ der}$, se rechaza la hipótesis fundamental.

§ 13. Comparación de la media muestral y la media general hipotética de un conjunto normal

A. La dispersión del conjunto general es conocida. Supongamos que el conjunto general X está distribuido normalmente, además, a pesar de que la media general a es desconocida existen motivos para suponer que ella es igual al valor hipotético (supuesto) a_0. Por ejemplo, si X es el conjunto de dimensiones x_i de una partida de piezas producidas por una máquina automática, se puede suponer que la media general a de estas dimensiones es igual a la dimensión teórica a_0. Para verificar esta suposición se halla la media muestral $\overline{x}$ y se establece si $\overline{x}$ y a_0 se diferencian de manera significativa o insignificativa. Si la diferencia resulta insignificativa, la máquina asegura en término medio la dimensión de proyecto; si la diferencia es significativa, la máquina debe reajustarse.

Supongamos que conocemos la dispersión del conjunto general, por ejemplo, del experimento anterior, o bien se ha hallado teóricamente, o bien se calculó por una muestra de gran volumen (por una muestra grande puede obtenerse una estimación bastante buena de la dispersión).

Sea así que de un conjunto general normal se ha extraído una muestra de volumen n y por ella se ha encontrado la media muestral $\overline{x}$ y al mismo tiempo la dispersión general σ^2 es conocida. Para un nivel de significación dado, se necesita

verificar por la media muestral la hipótesis fundamental H_0: $a = a_0$ de la igualdad entre la media general y el valor hipotético a_0.

Teniendo en cuenta que la media muestral es la estimación no desviada de la media general (cap. XVI, § 5), es decir, $M(\overline{X}) = a$, la hipótesis fundamental puede escribirse así: $M(\overline{X}) = a_0$.

Por consiguiente, se necesita verificar que la esperanza matemática de la media muestral es igual a la media general hipotética. En otras palabras, hay que establecer si las medias muestral y general se diferencian de manera significativa o insignificativa.

Como criterio de verificación de la hipótesis fundamental tomamos la magnitud aleatoria

$$U = \frac{\overline{X} - a_0}{\sigma(\overline{X})} = \frac{(\overline{X} - a_0)\sqrt{n}}{\sigma},$$

que está distribuida normalmente; además, cuando la hipótesis fundamental es cierta, $M(U) = 0$, $\sigma(U) = 1$.

Puesto que aquí la región crítica se construye en función del tipo de hipótesis alternativa, como en el § 10, nos limitaremos a formular la regla de verificación de la hipótesis fundamental, designando el valor del criterio U, calculado por los datos de las observaciones, por U_{obs}.

Regla 1. Para verificar, a un nivel de significación dado, la hipótesis fundamental H_0: $a = a_0$ de la igualdad entre la media general a de un conjunto normal con dispersión conocida σ^2 y el valor hipotético a_0, cuando la hipótesis alternativa H_1: $a \neq a_0$, hay que calcular el valor observado del criterio

$$U_{obs} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$$

y por la tabla de la función de Laplace, hallar el punto crítico de la región crítica bilateral por la igualdad

$$\Phi(u_{cr}) = \frac{1 - \alpha}{2}.$$

Si $|U_{obs}| < u_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $|U_{obs}| > u_{cr}$, la hipótesis fundamental se rechaza,

Regla 2. Cuando la hipótesis concurrente H_1: $a > a_0$, el punto crítico de la región crítica de derecha se halla por la igualdad

$$\Phi(u_{cr}) = \frac{1-2\alpha}{2}.$$

Si $U_{obs} < u_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $U_{obs} > u_{cr}$, la hipótesis fundamental se rechaza.

Regla 3. Para la hipótesis alternativa H_1: $a < a_0$, al principio se halla el punto crítico u_{cr} según la regla 2, y después se supone que el límite de la región crítica de izquierda es

$$u'_{cr} = -u_{cr}.$$

Si $U_{obs} > -u_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $U_{obs} < -u_{cr}$, se rechaza la hipótesis fundamental.

Ejemplo 1. De un conjunto general normal con desviación cuadrática media conocida $\sigma = 0{,}36$ se ha escogido una muestra de volumen $n = 36$ y por ella se ha hallado la media muestral $\bar{x} = 21{,}6$. Para el nivel de significación 0,05 hay que verificar la hipótesis fundamental H_0: $a = a_0 = 21$, siendo la hipótesis alternativa H_1: $a \neq 21$.

SOLUCION. Hallamos el valor observado del criterio

$$U_{obs} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = \frac{(21{,}6-21)\sqrt{36}}{0{,}36} = 10.$$

Por los datos, la hipótesis alternativa tiene la forma $a \neq a$, por eso la región crítica es bilateral.

Hallamos el punto crítico por la igualdad

$$\Phi(u_{cr}) = \frac{1-\alpha}{2} = \frac{1-0{,}5}{2} = 0{,}475.$$

Por la tabla de la función de Laplace encontramos $u_{cr} = 1{,}96$.

Puesto que $U_{obs} > u_{cr}$, rechazamos la hipotesis fundamental. En otras palabras, las medias muestral y general hipotética se diferencian de manera significativa.

Ejemplo 2. Por los datos del ejemplo 1 verificar la hipótesis fundamental H_0: $a = 21$, cuando la hipótesis alternativa es $a > 21$.

SOLUCION. Dado que la hipótesis alternativa tiene la forma $a > 21$, la región crítica es de derecha.

Hallamos el punto crítico por la igualdad

$$\Phi(u_{cr}) = \frac{1-2\alpha}{2} = \frac{1-2\cdot 0,05}{2} = 0,45.$$

Por la tabla de la función de Laplace encontramos $u_{cr} = 1,65$.

En virtud de que $U_{obs} = 10 > u_{cr}$, rechazamos la hipótesis fundamental; la diferencia entre las medias muestral y general hipotética es significativa.

Cabe hacer notar que en el ejemplo 2 la hipótesis fundamental podía haberse rechazado inmediatamente, ya que había sido rechazada en el ejemplo 1, para la región crítica bilateral. Hemos expuesto la resolución detallada con fines didácticos.

B. La dispersión del conjunto general es desconocida. Si la dispersión del conjunto general no es conocida (por ejemplo, en el caso de muestras pequeñas), como criterio de verificación de la hipótesis fundamental se toma la magnitud aleatoria

$$T = \frac{(\bar{X} - a_0)\sqrt{n}}{s},$$

donde s es la desviación cuadrática media «corregida». La magnitud T tiene la distribución t de Student con $k = n - 1$ grados de libertad.

La región crítica se construye en función del tipo de hipótesis alternativa. Ya que esto se realiza como se describió antes, nos limitamos a las reglas de verificación de la hipótesis fundamental.

Regla 1. Para un nivel de significación dado α, a fin de verificar la hipótesis fundamental H_0: $a = a_0$ sobre la igualdad entre la media general a desconocida (conjunto normal con dispersión desconocida) y el valor hipotético a_0, cuando la hipótesis alternativa H_1: $a \neq a_0$, hay que calcular el valor observado del criterio:

$$T_{obs} = \frac{(\bar{x} - a_0)\sqrt{n}}{s}$$

y por la tabla de puntos críticos de la distribución t de Student, según el nivel de significación dado α, situado en la línea superior de la tabla, y el número de grados de libertad $k = n - 1$, hallar el punto crítico $t_{cr.\ bilat}(\alpha, k)$.

Si $|T_{obs}| < t_{cr.\ bilat}$, no hay porque rechazar la hipótesis fundamental.

Si $|T_{obs}| > t_{cr.\,bilat}$, la hipótesis fundamental se rechaza.

Regla 2. Si la hipótesis alternativa $H_1 : a > a_0$, por el nivel de significación α, situado en la línea inferior de la tabla (suplemento 6), y el número de grados de libertad $k = n - 1$, se halla el punto crítico $t_{cr.\,der}(\alpha, k)$ de la región crítica de derecha.

Si $T_{obs} < t_{cr.\,der}$, no hay porque rechazar la hipótesis fundamental.

Si $T_{obs} > t_{cr.\,der}$, la hipótesis fundamental se rechaza.

Regla 3. Cuando la hipótesis alternativa $H_1 : a < a_0$, al principio se halla el punto crítico "auxiliar" $t_{cr.\,der}(\alpha, k)$ y se supone el límite de la región crítica de izquierda $t_{cr.\,izq} = -t_{cr.\,der}$.

Si $T_{obs} > -t_{cr.\,der}$, no hay porque rechazar la hipótesis fundamental.

Si $T_{obs} < -t_{cr.\,der}$, la hipótesis fundamental se rechaza.

Ejemplo 3. Por una muestra de volumen $n = 20$, escogida de un conjunto general normal, se ha hallado la media muestral $\overline{x} = 16$ y la desviación cuadrática media «corregida» $s = 4{,}5$. Para el nivel de significación 0,05 hay que verificar la hipótesis fundamental $H_0 : a = a_0 = 15$, cuando la hipótesis alternativa es $H_1 : a \neq 15$.

SOLUCIÓN. Calculamos el valor observado del criterio

$$T_{obs} = \frac{(\overline{x} - a_0)\sqrt{n}}{s} = \frac{(16 - 15)\cdot\sqrt{20}}{4{,}5} = 0{,}99.$$

Por los datos, la hipótesis alternativa tiene la forma $a \neq a_0$, por eso la región crítica es bilateral.

Por la tabla de puntos críticos de la distribución t de Student, según el nivel de significación $\alpha = 0{,}05$, situado en la línea superior de la tabla, y por el número de grados de libertad $k = 20 - 1 = 19$, hallamos el punto crítico $t_{cr.\,bilat}(0{,}05;\ 19) = 2{,}09$.

Puesto que $|T_{obs}| < t_{cr.\,bilat}$, no hay porque rechazar la hipótesis fundamental; la media muestral se diferencia de manera insignificativa de la media general hipotética.

§ 14. Vínculo entre la región crítica bilateral y el intervalo de confianza

Se demuestra fácilmente que al buscar la región crítica bilateral para el nivel de significación α, por lo tanto, se halla también el correspondiente intervalo de confianza

(o intervalo confidencial) con fiabilidad $\gamma = 1 - \alpha$. Por ejemplo, en el § 13, al verificar la hipótesis fundamental H_0: $a = a_0$, cuando H_1: $a \neq a_0$, exigimos que la probabilidad de caer el criterio $U = \frac{(\bar{x} - a)\sqrt{n}}{\sigma}$ en la región crítica bilateral fuese igual al nivel de significación α, por lo tanto, la probabilidad de que el criterio caiga en la región de aceptación de la hipótesis $(-u_{cr}, u_{cr})$, igual a $1 - \alpha = \gamma$. En otras palabras, con la fiabilidad γ se cumple la desigualdad

$$-u_{cr} < \frac{(\bar{x} - a)\sqrt{n}}{\sigma} < u_{cr},$$

o bien la desigualdad equivalente

$$\bar{x} - u_{cr}\frac{\sigma}{\sqrt{n}} < a < \bar{x} + u_{cr}\frac{\sigma}{\sqrt{n}}, \qquad (*)$$

donde $\Phi(u_{cr}) = \frac{\gamma}{2}$.

Hemos obtenido el intervalo de confianza para estimar la esperanza matemática a de una distribución normal en el caso de σ conocido, con fiabilidad γ (cap. XVI, § 15).

Nota. A pesar de que la búsqueda de la región crítica bilateral y del intervalo confidencial nos conduce a idénticos resultados, su interpretación es distinta: la región crítica bilateral define los límites (puntos críticos), entre los cuales está situado el $(1 - \alpha)$ % de número de criterios observados, hallados al repetir los experimentos; el intervalo confidencial determina los límites (extremos del intervalo), entre los cuales en $\gamma = (1 - \alpha)$% de pruebas está encerrado el valor verdadero del parámetro que se estima.

§ 15. Determinación del volumen mínimo de una muestra al comparar las medias muestral y general hipotética

Frecuentemente en la práctica se conoce la magnitud (precisión) $\delta > 0$ que no debe ser mayor que la magnitud absoluta de la diferencia entre las medias muestral y general hipotética. Por ejemplo, generalmente se necesita que la dimensión media de las piezas producidas se diferencia de la de proyecto no más que un δ prefijado.

Nos preguntamos: ¿cúal debe ser el volumen mínimo de la muestra para que se cumpla esta condición con la probabilidad $\gamma = 1 - \alpha$ (α es el nivel de significación)?

Puesto que el problema de encontrar el intervalo de confianza para estimar la esperanza matemática de una distri-

bución normal cuando se conoce σ y el problema de hallar la región crítica bilateral para verificar la hipótesis sobre la igualdad de la esperanza matemática (media general) al valor hipotético (§ 13, A) se reduce a lo mismo (§ 14), utilizamos la fórmula (cap. XVI, § 15)

$$n=\frac{u_{cr}^2\sigma^2}{\delta^2},$$

donde u_{cr} se halla por la igualdad $\Phi(u_{cr})=\frac{\gamma}{2}=\frac{1-\alpha}{2}$.

Si σ es deconocido, y se ha encontrado su estimación s, entonces (§ 13, B)

$$n=\frac{t_{cr.\,bilat}^2(\alpha,\,k)\cdot s^2}{\delta^2}.$$

§ 16. Ejemplo de hallazgo de la potencia del criterio

Mostremos un ejemplo de hallazgo de la potencia del criterio.

Ejemplo. Por la muestra de volumen $n=25$, escogida de un conjunto general normal con desviación cuadrática media $\sigma=10$ conocida, se ha hallado la media muestral $\bar{x}=18$. Para el nivel de significación 0,05 se necesita:

a) hallar la región crítica, si se verifica la hipótesis fundamental H_0: $a=a_0=20$ de la igualdad entre la media general y el valor hipotético, siendo la hipótesis alternativa $H_1: a<20$;

b) hallar la potencia del criterio de verificación, para $a_0=16$.

SOLUCIÓN. a) Puesto que la hipótesis alternativa tiene la forma $a<a_0$, la región crítica es de izquierda.

Utilizando la regla 3 (§ 13, A), hallamos el punto crítico: $u'_{cr}=-1{,}65$. Por lo tanto, la región crítica de izquierda se determina por la desigualdad $U<-1{,}65$, o desarrollado

$$\frac{(\bar{x}-20)\sqrt{25}}{10}<-1{,}65.$$

De aquí $\bar{x}<16{,}7$.

Para estos valores de la media muestral la hipótesis fundamental se rechaza; en este sentido $\bar{x}=16{,}7$ se puede considerar como valor crítico de la media muestral.

b) Para calcular la potencia del criterio examinado, encontramos previamente su valor, a condición de que la hipótesis alternativa sea cierta (es decir cuando $a_0 = 16$), poniendo $\bar{x} = 16,7$:

$$U = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = \frac{(16,7 - 16)\sqrt{25}}{10} = 0,35.$$

De aquí se aprecia que, si $\bar{x} < 16,7$, $U < 0,35$. Puesto que para $\bar{x} < 16,7$ la hipótesis fundamental se rechaza, y para $U < 0,35$ también se rechaza (en este caso la hipótesis alternativa es cierta, como supusimos $a_0 = 16$).

Utilizando la función de Laplace, a continuación hallamos la potencia del criterio, es decir, la probabilidad de que la hipótesis fundamental sea rechazada si la hipótesis alternativa es cierta (§ 7):

$$P(U < 0,35) = P(-\infty < U < 0,35) = P(-\infty < U < 0) + P(0 < U < 0,35) = 0,5 + \Phi(0,35) = 0,5 + 0,1368 = 0,6368.$$

Así pues, la potencia buscada del criterio examinado es aproximadamente igual a 0,64. Si se aumenta el volumen de la muestra, la potencia se incrementa.

Por ejemplo, cuando $n = 64$ la potencia es igual a 0,71. Si se aumenta α, la potencia también se incrementa. Por ejemplo, para $\alpha = 0,1$ la potencia es igual a 0,7642.

Nota. Conociendo la potencia se halla fácilmente la probabilidad del error de segundo género: $\beta = 1 - 0,64$. (Está claro que para resolver el ejemplo antes podía haberse encontrado β, y luego la potencia, igual a $1 - \beta$).

§ 17. Comparación de dos medias de conjuntos generales normales con dispersiones desconocidas (muestras dependientes)

En el párrafo anterior se supuso que las muestras eran independientes. Aquí se examinan las muestras de igual volumen, cuyas variantes son dependientes de dos en dos. Por ejemplo, si x_i ($i = 1, 2, \ldots, n$) son los resultados de las mediciones de las piezas por el primer aparato, e y_i son los resultados de las mediciones de las mismas piezas producidas en igual orden por un segundo aparato, x_i e y_i son dependientes de dos en dos y, bajo este concepto, las propias muestras son dependientes. Dado que, como regla $x_i \neq y_i$, se hace necesario establecer si los pares de estos números se diferencian de manera significativa o insignificativa.

Se plantea un problema análogo al comparar *dos* métodos de investigación realizados en *un mismo* laboratorio, o bien al investigar con *igual* método *en dos* laboratorios distintos.

De este modo, supongamos que los conjuntos generales X e Y están distribuidos normalmente; además sus dispersiones son desconocidas. Se necesita verificar, para el nivel de significación α, la hipótesis fundamental H_0: $M(\overline{X}) = M(\overline{Y})$ de la igualdad entre las medias generales de los conjuntos normales con dispersiones desconocidas, siendo la hipótesis alternativa H_1: $M(\overline{X}) \neq M(\overline{Y})$, por dos muestras dependientes de igual volumen.

Este problema de comparación de *dos* medias lo reducimos al problema de comparación de *una* media muestral con el valor hipotético de la media general, resuelto en § 13, B.

Con este propósito ponemos en examen las magnitudes aleatorias, o sea, las diferencias $D_i = X_i - Y_i$ y su media:

$$\overline{D} = \frac{\sum D_i}{n} = \frac{\sum (X_i - Y_i)}{n} = \frac{\sum X_i}{n} - \frac{\sum Y_i}{n} = \overline{X} - \overline{Y}.$$

Si la hipótesis nula es cierta, es decir, $M(\overline{X}) = M(\overline{Y})$, entonces $M(\overline{X}) - M(\overline{Y}) = 0$ y, por lo tanto,

$$M(\overline{D}) = M(\overline{X} - \overline{Y}) = M(\overline{X}) - M(\overline{Y}) = 0.$$

Por consiguiente, la hipótesis nula H_0: $M(\overline{X}) = M(\overline{Y})$ podemos escribirla así:

$$H_0\text{: } M(\overline{D}) = 0.$$

En este caso, la hipótesis alternativa toma la forma:

$$H_1\text{: } M(\overline{D}) \neq 0.$$

Nota 1. En adelante las diferencias no aleatorias observadas $x_i - y_i$ las designaremos por d_i, a distinción de las diferencias aleatorias $D_i = X_i - Y_i$. Análogamente la media muestral de estas diferencias $\frac{\sum d_i}{n}$ la designamos por $\overline{d}$, a distinción de la magnitud aleatoria $\overline{D}$.

Así pues, el problema de comparar las dos medias $\overline{x}$ e $\overline{y}$ se reduce a comparar una media muestral $\overline{d}$ con el valor

hipotético de la media general $M(\overline{D}) = a_0 = 0$. Este problema ya se resolvió en el § 13, B, por eso daremos sólo la regla de verificación de la hipótesis fundamental y un ejemplo que la ilustra.

Nota 2. Como se deduce de lo expuesto antes, en la fórmula (§ 13, B)

$$T_{obs} = \frac{(\overline{x} - a_0)\sqrt{n}}{s}$$

hay que poner

$$\overline{x} = \overline{d},\quad a_0 = 0,\quad s = s_d = \sqrt{\frac{\sum d_i^2 - \frac{\left[\sum d_i\right]^2}{n}}{n-1}}.$$

Luego $T_{obs} = \frac{\overline{d}\sqrt{n}}{s_d}$.

Regla. Para verificar, a un nivel de significación prefijado α, la hipótesis fundamental H_0: $M(\overline{X}) = M(\overline{Y})$ de la igualdad entre dos medias de conjuntos normales con dispersiones desconocidas (en el caso de muestras dependientes de igual volumen) cuando la hipótesis alternativa es $M(\overline{X}) \neq M(\overline{Y})$, hay que calcular el valor observado del criterio

$$T_{obs} = \frac{\overline{d}\sqrt{n}}{s_d}$$

y por la tabla de puntos críticos de la distribución t de Student, según el nivel de significación α, situado en la línea superior de la tabla y por el número de grados de libertad $k = n - 1$, hallar el punto crítico $t_{cr.\,bilat}(\alpha, k)$.

Si $|T_{obs}| < t_{cr.\,bilat}$, no hay porque rechazar la hipótesis fundamental.

Si $|T_{obs}| > t_{cr.\,bilat}$, la hipótesis fundamental se rechaza.

Ejemplo. 5 piezas se han medido en igual orden empleando dos aparatos y se han obtenido los siguientes resultados (en centésimas de mm):

$$x_1 = 6,\quad x_2 = 7,\quad x_3 = 8,\quad x_4 = 5,\quad x_5 = 7;$$
$$y_1 = 7,\quad y_2 = 6,\quad y_3 = 8,\quad y_4 = 7,\quad y_5 = 8.$$

Para el nivel de significación 0,05 hay que establecer si los resultados de las mediciones se diferencian de manera significativa o insignificativa.

SOLUCIÓN. Restando de los números de la primera línea los números de la segunda, obtenemos

$$d_1 = -1,\ d_2 = 1,\ d_3 = 0,\ d_4 = -2,\ d_5 = -1.$$

Hallamos la media muestral:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-1+1+0-2+(-1)}{5} = -0{,}6.$$

Teniendo en cuenta que $\sum d_i^2 = 1+1+4+1=7$ y $\sum d_i = -3$, hallamos la desviación cuadrática media «corregida»:

$$s_d = \sqrt{\frac{\sum d_i^2 - \frac{\left[\sum d_i\right]^2}{n}}{n-1}} = \sqrt{\frac{7-\frac{9}{5}}{5-1}} = \sqrt{1{,}3}.$$

Calculamos el valor observado del criterio

$$T_{\text{obs}} = \frac{\bar{d}\sqrt{n}}{s_d} = -\frac{0{,}6\sqrt{5}}{\sqrt{1{,}3}} = -1{,}18.$$

Por la tabla de puntos críticos de la distribución t de Student, según el nivel de significación 0,05, situado en la línea superior de la tabla, y el número de grados de libertad $k = 5-1=4$, hallamos el punto crítico $t_{\text{cr. bilat}}(0{,}05;\ 4) = 2{,}78$.

Puesto que $|T_{\text{obs}}| < t_{\text{cr. bilat}}$, no hay porque rechazar la hipótesis fundamental. En otras palabras, los resultados de las mediciones se diferencian de manera insignificativa.

§ 18. Comparación de la frecuencia relativa observada con la probabilidad hipotética de aparición de un suceso

Supongamos que por un número suficientemente grande n de experimentos independientes, en cada uno de los cuales la probabilidad p de que aparezca el suceso es constante, pero desconocida, se encontró la frecuencia relativa $\frac{m}{n}$. Admitamos que existen razones para suponer que la probabilidad incógnita es igual a un valor hipotético p_0. Para un nivel de significación dado α, hay que verificar la hipótesis nula consistente en que la probabilidad desconocida p es igual a la probabilidad hipotética p_0.

Dado que la probabilidad se estima por la frecuencia relativa, el problema examinado puede formularse, también, así: hay que establecer si la frecuencia relativa observada y la probabilidad hipotética se diferencian de manera significativa o insignificativa.

Como criterio de verificación de la hipótesis fundamental tomamos la magnitud aleatoria

$$U=\frac{\left(\frac{M}{n}-p_0\right)\sqrt{n}}{\sqrt{p_0q_0}},$$

donde $q_0 = 1 - p_0$.

La magnitud U, cuando la hipótesis fundamental es cierta, está distribuida con aproximación normalmente con parámetros $M(U) = 0$, $\sigma(U) = 1$.

Explicación. Se ha demostrado (teorema de Laplace) que para valores bastante grandes de n la frecuencia relativa tiene aproximadamente una distribución normal con esperanza matemática p y desviación cuadrática media $\sqrt{\frac{pq}{n}}$. Normando la frecuencia relativa (restando la esperanza matemática y dividiendo por la desviación cuadrática media) obtenemos

$$U=\frac{\frac{M}{n}-p}{\sqrt{\frac{pq}{n}}}=\frac{\left(\frac{M}{n}-p\right)\sqrt{n}}{\sqrt{pq}},$$

además $M(U) = 0$, $\sigma(U) = 1$.

Si la hipótesis nula es válida, es decir, para $p = p_0$,

$$U=\frac{\left(\frac{M}{n}-p_0\right)\sqrt{n}}{\sqrt{p_0q_0}}.$$

Nota 1. En adelante la frecuencia observada se designará por $\frac{m}{n}$ a diferencia de la magnitud aleatoria $\frac{M}{n}$

Puesto que aquí la región crítica se construye como en el § 10, expondremos solamente las reglas de verificación de la hipótesis nula y un ejemplo ilustrativo.

Regla 1. Para verificar, a un nivel de significación dado, la hipótesis fundamental H_0: $p = p_0$ de la igualdad entre

la probabilidad incógnita y la probabilidad hipotética, siendo la hipótesis alternativa H_1: $p \neq p_0$, hay que calcular el valor observado del criterio

$$U_{obs} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}$$

y por la tabla de la función de Laplace hallar el punto crítico u_{cr} por la igualdad $\Phi(u_{cr}) = \frac{1-\alpha}{2}$.

Si $|U_{obs}| < u_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $|U_{obs}| > u_{cr}$, la hipótesis nula se rechaza.

Regla 2. Para la hipótesis alternativa $H_1: p > p_0$ se halla el punto crítico de la región crítica de derecha por la igualdad $\Phi(u_{cr}) = \frac{1-2\alpha}{2}$.

Si $U_{obs} < u_{cr}$, no hay porque rechazar la hipótesis nula.

Si $U_{obs} > u_{cr}$, la hipótesis nula se rechaza.

Regla 3. Para la hipótesis alternativa H_1: $p < p_0$ se halla el punto crítico u_{cr} por la regla 2, y luego se supone el límite de la región crítica de izquierda $u'_{cr} = -u_{cr}$.

Si $U_{obs} > -u_{cr}$, no hay porque rechazar la hipótesis cero.

Si $U_{obs} < -u_{cr}$, la hipótesis nula se rechaza.

Nota 2. Los resultados satisfactorios asegura el cumplimiento de la desigualdad $np_0q_0 > 9$.

Ejemplo. Por 100 pruebas independientes se ha determinado la frecuencia relativa 0,08. Verificar la hipótesis cero H_0: $p = p_0 = 0{,}12$ para el nivel de significación 0,05, siendo la hipótesis alternativa H_1: $p \neq 0{,}12$.

SOLUCIÓN. Hallamos el valor observado del criterio

$$U_{obs} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}} = \frac{(0{,}08 - 0{,}12)\sqrt{100}}{\sqrt{0{,}12 \cdot 0{,}88}} = -1{,}23.$$

Según los datos, la hipótesis alternativa tiene la forma $p \neq p_0$, por eso la región crítica es bilateral.

Hallamos el punto crítico de la igualdad.

$$\Phi(u_{cr}) = \frac{1-\alpha}{2} = \frac{1-0{,}05}{2} = 0{,}475.$$

Por la tabla de la función de Laplace (suplemento 2) hallamos $u_{cr} = 1,96$.

Puesto que $|U_{obs}| < u_{cr}$, no hay porque rechazar la hipótesis fundamental. En otras palabras, la frecuencia relativa observada se diferencia de manera insignificativa de la probabilidad hipotética.

§ 19. Comparación de varias dispersiones de conjuntos generales normales por muestras de distinto volumen. Criterio de Bartlett

Supongamos que los conjuntos generales $X_1, X_2, \ldots, X_l$ están distribuidos normalmente. De estos conjuntos se han extraído muestras independientes, en general, de distintos volúmenes $n_1, n_2, \ldots, n_l$ (algunos volúmenes pueden ser idénticos; si todas las muestras tienen igual volumen, conviene utilizar el criterio de Cochran descripto en el párrafo siguiente). Por las muestras han hallado las dispersiones muestrales $s_1^2, s_2^2, \ldots, s_l^2$.

Para un nivel de significación dado α, se necesita verificar por las dispersiones muestrales corregidas la hipótesis fundamental consistente en que las dispersiones generales de los conjuntos examinados son iguales entre sí:

$$H_0: D(X_1) = D(X_2) = \ldots = D(X_l).$$

En otras palabras, hay que establecer si las dispersiones muestrales corregidas se diferencian de manera significativa o insignificativa.

La hipótesis sobre la igualdad de varias dispersiones aquí considerada se llama *hipótesis de homogeneidad de las dispersiones*.

Notemos que como número de grados de libertad de la dispersión s_i^2 se llama el número $k_i = n_i - 1$, es decir, el número menor en una unidad del volumen de la muestra, por la que se ha calculado la dispersión.

Designamos por $\overline{s^2}$ la media aritmética de las dispersiones corregidas, ponderada por los números de grados de libertad:

$$\overline{s^2} = \frac{\sum_{i=1}^{l} k_i s_i^2}{k},$$

donde $k = \sum_{i=1}^{l} k_i$

Como criterio de verificación de la hipótesis nula de homogeniedad de las dispersiones tomamos el criterio de Bartlett, o sea, la magnitud aleatoria

$$B = \frac{V}{C},$$

donde $V = 2{,}303[k \cdot \lg \overline{s^2} - \sum_{i=1}^{l} k_i \lg s_i^2]$,

$$C = 1 + \frac{1}{3(l-1)}\left[\sum_{i=1}^{l} \frac{1}{k_i} - \frac{1}{k}\right].$$

Bartlett estableció que la magnitud aleatoria B, a condición de que la hipótesis nula sea justa, está distribuida aproximadamente como χ^2 con $l - 1$ grados de libertad, si todo $k_i > 2$. Teniendo en cuenta que $k_i = n_i - 1$, deducimos que $n_i - 1 > 2$, o bien $n_i > 3$, es decir, el volumen de cada una de las muestras no debe ser mayor que 4.

La región crítica se construye de derecha, partiendo de la condición de que la probabilidad de caer el criterio en esta región suponiendo que la hipótesis fundamental es cierta, sea igual al nivel de significación admitido:

$$P[B > \chi^2_{cr}(\alpha, l-1)] = \alpha.$$

El punto crítico $\chi^2_{cr}(\alpha, l-1)$ se halla por la tabla (suplemento 5) según el nivel de significación α y el número de grados de libertad $k = l - 1$ luego de lo cual se determina la región crítica de derecha por la desigualdad

$$B > \chi^2_{cr},$$

en tanto que la región de aceptación de la hipótesis, por la desigualdad

$$B < \chi^2_{cr}.$$

Designamos el valor del criterio de Bartlett, calculado por los datos de las observaciones, por B_{obs} y formulamos la regla de verificación de la hipótesis nula.

Regla. Para verificar, a un nivel de significación prefijado α, la hipótesis nula sobre la homogeneidad de las dispersiones de conjuntos normales, hay que calcular el valor observado del criterio de Bartlett $B = \frac{V}{C}$ y por la tabla de puntos críticos de la distribución χ^2 encontrar el punto crítico $\chi^2_{cr}(\alpha, l-1)$.

Si $B_{obs} < \chi^2_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $B_{obs} > \chi^2_{cr}$, la hipótesis fundamental se rechaza.

Nota 1. No hay que apresurarse en calcular la constante C. Al principio hay que hallar V y compararla con χ^2_{cr}; si resulta que $V < \chi^2_{cr}$, tanto mejor (ya que $C > 1$) $B = \frac{V}{C} < \chi^2_{cr}$ y, por lo tanto, no hay porque calcular C.

Si $V > \chi^2_{cr}$, hay que calcular C y luego comparar B con χ^2_{cr}.

Nota 2. El criterio de Bartlett es muy sensible a las desviaciones de las distribuciones respecto de la normal, por eso es necesario tratar con cuidado las deducciones obtenidas por este criterio.

Ejemplo. Por cuatro muestras independientes de volúmenes $n_1 = 10$, $n_2 = 12$, $n_3 = 15$, $n_4 = 16$, escogidas de conjuntos generales normales, se han encontrado las dispersiones muestrales corregidas, respectivamente iguales a 0,25; 0,40; 0,36; 0,46. Para el nivel de significación 0,05, verificar la hipótesis de homogenidad de las dispersiones (la región crítica es de derecha).

SOLUCIÓN. Formamos la tabla de cálculo 25 (la columna 8 por ahora no la completaremos, ya que aún no sabemos si se necesita calcular C):

Tabla 25

1	2	3	4	5	6	7	8
Número de muestra i	Volumen de la muestra n_i	Número de grados de libertad k_i	Dispersión s_i^2	$k_i s_i^2$	$\lg s_i^2$	$k_i \lg s_i^2$	$\frac{1}{k_i}$
1	10	9	0,25	2,25	$\bar{1}$,3979	$\bar{6}$,5811	
2	13	12	0,40	4,80	$\bar{1}$,6021	$\bar{5}$,2252	
3	15	14	0,36	5,04	$\bar{1}$,5563	$\bar{7}$,7822	
4	16	15	0,46	6,90	$\bar{1}$,6628	$\bar{6}$,9420	
$\sum$		$k = 50$		18,99		$\overline{22}$,5305	

Utilizando la tabla de cálculo, hallamos:

$$\overline{s^2} = \frac{\sum k_i s_i^2}{k} = \frac{18{,}99}{50} = 0{,}3798; \quad \lg 0{,}3798 = \bar{1}{,}5795;$$

$$V = 2{,}303\,[k \lg \overline{s^2} - \sum k_i \lg s_i^2] =$$
$$= 2{,}303\,[50 \cdot \bar{1}{,}5795 - \overline{22}{,}5305] = 1{,}02.$$

Por la tabla (suplemento 5), según el nivel de significación 0,05 y el número de grados de libertad $l - 1 = 4 - 1 = 3$, hallamos el punto crítico χ^2_{cr} (0,05; 3) = 7,8.

Puesto que $V < \chi^2_{cr}$, tanto mejor (puesto que $C > 1$) $B_{obs} = \frac{V}{C} < \chi^2_{cr}$ y, por lo tanto, no hay porque rechazar la hipótesis cero sobre la homogeneidad de las dispersiones. En otras palabras, las dispersiones muestrales corregidas se diferencian de manera insignificativa.

Nota 3. Si hay que estimar la dispersión general, entonces para la condición de homogeneidad de las dispersiones, conviene tomar como su estimación la media aritmética de las dispersiones corregidas, ponderada por los números de grados de libertad, es decir

$$\overline{s^2} = \frac{\sum k_i s_i^2}{k}.$$

Por ejemplo, en el problema examinado como estimación de la dispersión general conviene tomar 0,3798.

§ 20. Comparación de varias dispersiones de conjuntos generales normales por muestras de igual volumen. Criterio de Cochran

Supongamos que los conjuntos generales $X_1, X_2, \ldots, X_l$ están distribuidos normalmente. De estos conjuntos se han escogido l muestras independientes de *igual volumen* n y por ellas se han encontrado las dispersiones muestrales corregidas $s_1^2, s_2^2, \ldots, s_l^2$, todo con igual número de grados de libertad $k = n - 1$.

Por las dispersiones corregidas, para un nivel de significación prefijado α, se necesita verificar la hipótesis fundamental consistente en que las dispersiones generales de los conjuntos examinados son iguales entre sí:

$$H_0:\ D(X_1) = D(X_2) = \ldots = D(X_l).$$

En otras palabras, hay que verificar: las dispersiones muestrales corregidas se diferencian de manera significativa o insignificativa.

En el caso examinado de muestras de igual volumen, según el criterio de Fisher-Snedecor (§ 8) es posible comparar las dispersiones máxima y mínima; si resulta que la diferencia entre ellas es insignificativa, también es insignificativa la diferencia entre las demás dispersiones. El inconveniente de este método está en que la información que contiene las restantes dispersiones, salvo la mínima y la máxima, no se tomará en consideración.

También se puede aplicar el criterio de Bartlett. Sin embargo, como se indicó en el § 19, es conocida sólo la distribución *aproximada* de este criterio, por eso conviene utilizar el criterio de Cochran, cuya distribución se ha encontrado *con exactitud*.

De este modo, como criterio de verificación de la hipótesis nula tomamos el criterio de Cochran, es decir, la relación entre la dispersión máxima corregida y la suma de todas las dispersiones corregidas:

$$G = \frac{s^2_{máx}}{s_1^2 + s_2^2 + \ldots + s_l^2}.$$

La distribución de esta magnitud aleatoria depende sólo del número de grados de libertad $k = n - 1$ y la cantidad de muestras l.

La región crítica se construye de derecha, partiendo de la condición de que la probabilidad de que el criterio caiga en esta región, suponiendo que la hipótesis nula es cierta, sea igual al nivel de significación admitido

$$P[G > G_{cr}(\alpha, k, l)] = \alpha.$$

El punto crítico $G_{cr}(\alpha, k, l)$ se halla por la tabla (suplemento 8) y entonces, la región crítica de derecha se determina por la desigualdad

$$G > G_{cr},$$

y la región de aceptación de la hipótesis nula, por la desigualdad

$$G < G_{cr}.$$

El valor del criterio, calculado por los datos de las observaciones, lo designamos por G_{obs} y formulamos la regla de verificación de la hipótesis nula.

Regla. Para verificar, a un nivel de significación dado α, la hipótesis de homogeneidad de las dispersiones de conjuntos distribuidos normalmente, hay que calcular el valor observado del criterio y por la tabla hallar el punto crítico.

Si $G_{obs} < G_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $G_{obs} > G_{cr}$, la hipótesis fundamental se rechaza.

Nota. Si hay que estimar la dispersión general, entonces a condición de la homogeneidad de las dispersiones, como estimación de la misma conviene tomar la media aritmética de las dispersiones muestrales corregidas.

Ejemplo. Por cuatro muestras independientes de igual volumen $n = 17$, escogidas de conjuntos generales normales, se han encontrado las dispersiones corregidas: 0,26; 0,36; 0,40; 0,42. Se necesita: a) para el nivel de significación 0,05, verificar la hipótesis nula (fundamental) sobre la homogeneidad de las dispersiones generales (la región crítica es de derecha), b) estimar la dispersión general.

SOLUCION. a) Hallamos el valor observado del criterio de Cochran, es decir, la relación entre la dispersión corregida máxima y la suma de todas las dispersiones:

$$G_{obs} = \frac{0,42}{0,26+0,36+0,40+0,42} = 0,2917.$$

Según la tabla (suplemento 8), el nivel de significación 0,05, el número de grados de libertad $k = 17 - 1 = 16$ y el número de muestras $l = 4$, hallamos el punto crítico $G_{cr}(0,05;\ 16;\ 4) = 0,4366$.

Puesto que $G_{obs} < G_{cr}$, no hay porque rechazar la hipótesis cero sobre la homogeneidad de las dispersiones. En otras palabras, las dispersiones muestrales corregidas se diferencian de manera insignificativa.

b) Dado que la hipótesis fundamental es válida, como estimación de la dispersión general tomamos la media aritmética de las dispersiones corregidas:

$$\sigma^2 = \frac{0,26+0,36+0,40+0,42}{4} = 0,36.$$

§ 21. Verificación de la hipótesis de significación del coeficiente de correlación muestral

Supongamos que el conjunto general bidimensional (X, Y) está distribuido normalmente. De este conjunto se ha extraído una muestra de volumen n y por ella se ha encontrado el coeficiente de correlación muestral r_m, que resultó distinto de cero. Puesto que la muestra se ha escogido de manera aleatoria, aún no se puede deducir que el coeficiente de correlación del conjunto general r_g también es distinto de cero. En resumidas cuentas, nos interesa precisamente este coeficiente, por eso se hace necesario, para un nivel de significación dado α, verificar la hipótesis fundamental H_0: $r_g = 0$ sobre la igualdad a cero del coeficiente de correlación general, siendo la hipótesis alternativa H_1: $r_g \neq 0$.

Si la hipótesis fundamental se rechaza, significa que el coeficiente de correlación muestral se diferencia de cero de manera significativa, mientras que las X e Y están correlacionadas, es decir, están vinculadas por una dependencia lineal.

Si la hipótesis fundamental se admita, el coeficiente de correlación muestral es insignificativa, y las X e Y no están correlacionadas, es decir no están vinculadas por una dependencia lineal.

Como criterio de verificación de la hipótesis nula tomamos la magnitud aleatoria

$$T = \frac{r_m \sqrt{n-2}}{\sqrt{1-r_m^2}}.$$

La magnitud T tiene distribución t de Student con $k = n - 2$ grados de libertad, si la hipótesis fundamental es válida.

Puesto que la hipótesis alternativa tiene la forma $r_g \neq 0$, la región crítica es bilateral; ésta se construye igual que en el § 12 (primer caso).

El valor del criterio, calculado por los datos de las observaciones, lo designamos por T_{obs} y formulamos la regla de verificación de la hipótesis nula.

Regla. Para verificar, a un nivel de significación prefijado α, la hipótesis fundamental H_0: $r_g = 0$ sobre la igualdad a cero del coeficiente de correlación general de una magnitud aleatoria normal bidimensional, en el caso de la hipótesis alternativa, H_1: $r_g \neq 0$, hay que calcular el

valor observado del criterio

$$T_{obs}=\frac{r_m\sqrt{n-2}}{\sqrt{1-r_m^2}}$$

y por la tabla de puntos críticos de la distribución t de Student, según el nivel de significación dado y el número de grados de libertad $k=n-2$, hallar el punto crítico $t_{cr}(\alpha, k)$ para la región crítica bilateral.

Si $|T_{obs}|<t_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $|T_{obs}|>t_{cr}$, la hipótesis fundamental se rechaza.

Ejemplo. Por una muestra de volumen $n=122$, escogida de un conjunto normal bidimensional (X, Y), se encontró el coeficiente de correlación muestral $r_m=0,4$. Para el nivel de significación 0,05, verificar la hipótesis nula sobre la igualdad a cero del coeficiente de correlación general siendo la hipótesis alternativa $H_1: r_g\neq 0$.

SOLUCION. Hallamos el valor observado del criterio

$$T_{obs}=\frac{r_m\sqrt{n-2}}{\sqrt{1-r_m^2}}=\frac{0,4\sqrt{122-2}}{\sqrt{1-0,4^2}}=4,78.$$

Según los datos, la hipótesis alternativa tiene la forma $r_g\neq 0$, por eso, la región crítica es bilateral.

Por el nivel de significación 0,05 y el número de grados de libertad $k=122-2=120$, hallamos por la tabla (suplemento 6), para la región crítica bilateral, el punto crítico $t_{cr}(0,05;\ 120)=1,98$.

Ya que $T_{obs}>t_{cr}$, rechazamos la hipótesis fundamental. En otras palabras, el coeficiente de correlación muestral se diferencia de cero de manera significativa, es decir, X e Y están correlacionadas.

§ 22. Verificación de la hipótesis de distribución normal de un conjunto general. Criterio de aceptación de Pearson

En los párrafos anteriores supusimos conocida la ley de distribución del conjunto general.

Si la ley de distribución es desconocida, pero hay motivos de suponer que ella tiene una forma determinada (la llamamos A), entonces se verifica la hipótesis fundamental: el conjunto general está distribuido por la ley A.

La hipótesis sobre la supuesta ley de distribución desconocida se verifica de igual modo que para la hipótesis sobre los parámetros de la distribución, es decir, mediante una magnitud aleatoria especialmente escogida, o sea, el criterio de aceptación.

El criterio de verificación de la hipótesis sobre la supuesta ley de distribución desconocida se llama *criterio de aceptación*.

Existen varios criterios de aceptación de: χ^2 («ji cuadrado») K. de Pearson, Kolmogorov, Smirnov, etc.

Nos limitaremos a describir la aplicación del criterio de Pearson a la verificación de la hipótesis sobre la distribución normal de un conjunto general (el criterio se aplica análogamente para otras distribuciones, en esto reside su ventaja). Con este propósito vamos a comparar las frecuencias empíricas (observadas) y teóricas (calculadas, suponiendo una distribución normal).

Generalmente se distinguen las frecuencias empíricas y teóricas. Por ejemplo (cap. XVII, § 7).

frecuencias empír.	6	13	38	74	106	85	30	10	4
frecuencias teór.	3	14	42	82	99	76	37	11	2.

¿La divergencia de las frecuencias es aleatoria? Es posible que la divergencia sea aleatoria (insignificativa) y se debe al pequeño número de observaciones, o bien a su método de agrupamiento, o a otros motivos. Es posible que la divergencia de las frecuencias no es aleatoria (significativa) y se debe a que las frecuencias teóricas están calculadas, partiendo de la hipótesis sobre la distribución normal del conjunto general.

El criterio de Pearson responde a la pregunta antes formulada. En verdad, como cualquier criterio, no demuestra la certeza de la hipótesis, sino sólo establece, en el nivel de significación admitido, su acuerdo o no con los datos de las observaciones.

Así pues, supongamos que por la muestra de volumen n se ha obtenido la distribución empírica:

variantes	x_i	x_1	x_2	...	x_s,
frecuencias empír.	n_i	n_1	n_2	...	n_s.

Admitamos que en la hipótesis de distribución normal del conjunto general, se han calculado las frecuencias teóricas n_i'

(por ejemplo, como en el párrafo siguiente). Para el nivel de significación α, hay que verificar la hipótesis fundamental; el conjunto general está distribuido normalmente.

Como criterio de verificación de la hipótesis nula tomamos la magnitud aleatoria

$$\chi^2 = \sum \frac{(n_i - n_i')^2}{n_i'} . \qquad (*)$$

Esta magnitud es aleatoria, ya que en distintos experimentos ella toma diferentes valores, previamente desconocidos. Está claro que cuanto menos se diferencian las frecuencias empíricas y las teóricas, tanto menor es la magnitud del criterio (*) y, por lo tanto, ella caracteriza hasta cierto grado la proximidad de las distribuciones empírica y teórica.

Cabe hacer notar que al elevar al cuadrado las diferencias de frecuencias, se hace imposible la eliminación mutua de las diferencias positivas y negativas. Dividiendo por n_i' se logra reducir cada uno de los sumandos; en caso contrario la suma sería tan grande que daría lugar al rechazo de la hipótesis nula incluso cuando ella es válida. Desde luego, las consideraciones expuestas no son la argumentación del criterio elegido, sino solamente una aclaración.

Se ha demostrado que para $n \to \infty$ la ley de distribución de la magnitud aleatoria (*), independientemente de a qué ley de distribución obedece el conjunto general, tiende a la ley de distribución χ^2 con k grados de libertad. Por eso, la magnitud aleatoria (*) está designada por χ^2, y el propio criterio se llama criterio de aceptación «ji cuadrado».

El número de grados de libertad se halla por la igualdad
$k = s - 1 - r$,
donde s es el número de grupos (intervalos parciales) de la muestra;
r es el número de parámetros de la distribución supuesta que están estimados por los datos de la muestra.

En particular, si la distribución supuesta es normal, se estiman dos parámetros (la esperanza matemática y la desviación cuadrática media), por eso $r = 2$ y el número de grados de libertad $k = s - 1 - r = s - 1 - 2 = s - 3$.

Si, por ejemplo, se supone que el conjunto general está distribuido por la ley de Poisson, se estima un parámetro λ, por eso $r = 1$ y $k = s - 2$.

Ya que el criterio unilateral rechaza más «rigurosamente» la hipótesis nula que el bilateral, construimos la region crí-

tica de derecha, partiendo de la condición de que la probabilidad de caer el criterio en esta región, suponiendo que la hipótesis fundamental es cierta, sea igual al nivel de significación admitido α:

$$P\,[\chi^2 > \chi^2_{cr}(\alpha;\ k)] = \alpha.$$

Por consiguiente, la región crítica de derecha se determina por la desigualdad

$$\chi^2 > \chi^2_{cr}(\alpha;\ k),$$

y la región de aceptación de la hipótesis fundamental, por la desigualdad

$$\chi^2 < \chi^2_{cr}(\alpha;\ k).$$

El valor del criterio, calculado por los datos de las observaciones, lo designamos por χ^2_{cr} y formulamos la regla de verificación de la hipótesis fundamental.

Regla. Para verificar a un nivel de significación prefijado la hipótesis fundamental H_0: el conjunto general está distribuido normalmente, al principio hay que calcular las frecuencias teóricas, y después el valor observado del criterio

$$\chi^2_{obs} = \sum \frac{(n_i - n_i')^2}{n_i'} \qquad (**)$$

y por la tabla de puntos críticos de la distribución χ^2, según el nivel de significación dado α y el número de grados de libertad $k = s - 3$, hallar el punto crítico $\chi^2_{cr}(\alpha;\ k)$.

Si $\chi^2_{obs} < \chi^2_{cr}$, no hay porque rechazar la hipótesis fundamental.

Si $\chi^2_{obs} > \chi^2_{cr}$, la hipótesis fundamental se rechaza.

Nota 1. El volumen de la muestra debe ser bastante grande, en todo caso no menor que 50. Cada grupo debe contener no menos de 5—8 variantes; los grupos poco numerosos hay que unirlos en uno, sumando las frecuencias.

Nota 2. Puesto que son posibles errores de primer y segundo género, en particular, si la concordancia de las frecuencias teóricas y empíricas es «demasiado buena», hay que ser prudente. Por ejemplo, se puede repetir la prueba, aumentar el número de observaciones, utilizar otros criterios, construir la gráfica de distribución, calcular la asimetría y el exceso (cap. XVII, § 8).

Nota 3. A fines de controlar los cálculos, la fórmula (**) la transformamos a la forma

$$\chi^2_{obs} = \sum \frac{n_i^2}{n_i'} - n.$$

Recomendamos a los lectores realizar esta transformación individualmente, para los cuales necesario elevar al cuadrado en (**) la diferencia de frecuencias, simplificar el resultado por n'_i y tener en cuenta que $\sum n_i = n$, $\sum n'_i = n$.

Ejemplo. Verificar la hipótesis sobre la distribución normal de un conjunto general, para el nivel de significación 0,05, si son conocidas las frecuendias empíricas y teóricas:

frecuencias empír.	6	13	38	74	106	85	30	14
frecuencias teór.	3	14	42	82	99	76	37	13.

SOLUCION. Calculamos χ^2_{obs}, para ello formamos la tabla de cálculo 26.

Tabla 26

1	2	3	4	5	6	7	8
i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$	n_i^2	$\frac{n_i^2}{n'_i}$
1	6	3	3	9	3	36	12
2	13	14	−1	1	0,07	169	12,07
3	38	42	−4	16	0,38	1 444	34,38
4	74	82	−8	64	0,78	5 476	66,78
5	106	99	7	49	0,49	11 236	113,49
6	85	76	9	81	1,07	7 225	95,07
7	30	37	−7	49	1,32	900	24,32
8	14	13	1	1	0,08	196	15,08
$\sum$	366	366			$\chi^2_{obs} = 7,19$		373,19

Verificación: $\chi^2_{obs} = 7,19$:

$$\sum \frac{n_i^2}{n'_i} - n = 373,19 - 366 = 7,19.$$

Los cálculos son correctos.

Hallamos el número de grados de libertad, teniendo en cuenta que el número de grupos de la muestra (número de distintas variantes) $s = 8$; $k = 8 - 3 = 5$.

Por la tabla de puntos críticos de la distribución χ^2 (suplemento 5), según el nivel de significación $\alpha = 0,05$ y el número de grados de libertad $k = 5$, hallamos χ^2_{cr} (0,05; 5) $= 11,1$.

Puesto que $\chi^2_{obs} < \chi^2_{cr}$, no hay porque rechazar la hipótesis fundamental. En otras palabras, la divergencia entre las frecuencias empíricas y teóricas es insignificativa. Por lo tanto, los datos de las observaciones concuerdan con la hipótesis de distribución normal del conjunto general.

§ 23. Metodología del cálculo de las frecuencias teóricas de una distribución normal

Como se deduce del párrafo precedente, lo esencial del criterio de aceptación de Pearson es la comparación de las frecuencias empíricas y teóricas. Está claro que las frecuencias empíricas se hallan experimentalmente. ¿Cómo hallar las frecuencias teóricas, si se supone que el conjunto general está distribuido normalmente? A continuación se expone uno de los métodos de resolución de este problema.

1. Todo el intervalo de valores observados X (muestras de volumen n) se divide en s intervalos parciales (x_i, x_{i+1}) de igual longitud. Se encuentran los centros de los intervalos parciales $x_i^* = \frac{x_i + x_{i+1}}{2}$; como frecuencia n_i de las variantes x_i^* se toma el número de variantes que han caído en el i-ésimo intervalo. En suma se obtiene una sucesión de variantes equidistantes y sus correspondientes frecuencias:

$$x_1^* \quad x_2^* \ \ldots \ x_s^*,$$
$$n_1, \quad n_2 \ \ldots \ n_s,$$

además, $\Sigma n_i = n$.

2. Se calculan, por ejemplo, por el método de los productos, la media muestral $\overline{x}^*$ y la desviación cuadrática media muestral σ^*.

3. Se normaliza la magnitud aleatoria X, es decir, se pasa a la magnitud $Z = \frac{X - \overline{x}^*}{\sigma^*}$ y se calculan los extremos de los intervalos (z_i, z_{i+1}):

$$z_i = \frac{x_i - \overline{x}^*}{\sigma^*}, \quad z_{i+1} = \frac{x_{i+1} - \overline{x}^*}{\sigma^*},$$

además, el valor mínimo de Z, es decir, z_i se supone igual a $-\infty$, y el máximo, es decir, z_s se supone igual a ∞.

4. Se calculan las probabilidades teóricas p_i de que X caiga en los intervalos (x_i, x_{i+1}) por la igualdad [Φ (z), o sea,

la función de Laplace]

$$p_i = \Phi(z_{i+1}) - \Phi(z_i)$$

y, por último, se hallan las frecuencias teóricas $n'_i = np_i$.

Ejemplo. Hallar las frecuencias teóricas por una distribución de intervalo dado de la muestra de volumen $n = 200$, suponiendo que el conjunto general está distribuido normalmente (tabla 27).

Tabla 27

Número del intervalo	Límites del intervalo		Frecuencia	Número del intervalo	Límites del intervalo		Frecuencia
i	x_i	x_{i+1}	n_i	i	x_i	x_{i+1}	n_i
1	4	6	15	6	14	16	21
2	6	8	26	7	16	18	24
3	8	10	25	8	18	20	20
4	10	12	30	9	20	22	13
5	12	14	26				
							$n = 200$

SOLUCION 1. Hallamos los centros de los intervalos $x_i^* = \frac{x_i + x_{i+1}}{2}$. Por ejemplo, $x_1^* = \frac{4+6}{2} = 5$. De manera análoga obtenemos la sucesión de variantes equidistantes x_i^* y sus correspondientes frecuencias n_i:

$$\begin{array}{cccccccccc} x_i^* & 5 & 7 & 9 & 11 & 13 & 15 & 17 & 19 & 21 \\ n_i & 15 & 26 & 25 & 30 & 26 & 21 & 24 & 20 & 13. \end{array}$$

2. Utilizando el método de los productos, hallamos la media muestral y la disviación cuadrática media muestral:

$$\overline{x^*} = 12{,}63, \quad \sigma^* = 4{,}695.$$

3. Hallamos los intervalos (z_i, z_{i+1}), teniendo en cuenta que $\overline{x^*} = 12{,}63$, $\sigma^* = 4{,}695$, $\frac{1}{\sigma^*} = 0{,}213$, para lo cual componemos la tabla de cálculo 28.

4. Hallamos las probabilidades teóricas p_i y las frecuencias teóricas buscada $n'_i = np_i$, para lo cual componemos la tabla de cálculo 29.

Tabla 28

i	Límites del intervalo		$x_i-\overline{x^*}$	$x_{i+1}-\overline{x^*}$	Límites del intervalo	
	x_i	x_{i+1}			$z_i=\frac{x_i-\overline{x^*}}{\sigma^*}$	$z_{i+1}=\frac{x_{i+1}-\overline{x^*}}{\sigma^*}$
1	4	6	—	−6,63	$-\infty$	−1,41
2	6	8	−6,63	−4,63	−1,41	−0,99
3	8	10	−4,63	−2,63	−0,99	−0,56
4	10	12	−2,63	−0,63	−0,156	−0,13
5	12	14	−0,63	1,37	−0,13	0,29
6	14	16	1,37	3,37	0,29	0,72
7	16	18	3,37	5,37	0,72	1,14
8	18	20	5,37	7,37	1,14	1,57
9	20	22	7,37	—	1,57	∞

Las frecuencias teóricas buscadas se encuentran en la última columna de la tabla 29.

Tabla 29

i	Límites del intervalo		$\Phi(z_i)$	$\Phi(z_{i+1})$	$p_i=\Phi(z_{i+1})-\Phi(z_i)$	$n_i'=np_i=200\ p_i$
	z_i	z_{i+1}				
1	$-\infty$	−1,41	−0,5	−0,4207	0,0793	15,86
2	−1,41	−0,99	−0,4207	−0,3389	0,0818	16,36
3	−0,99	−0,56	−0,3389	−0,2123	0,1266	25,32
4	−0,56	−0,13	−0,2123	−0,0517	0,1606	32,16
5	−0,13	0,29	−0,0517	0,1141	0,1658	33,16
6	0,29	0,72	0,1141	0,2642	0,1501	30,02
7	0,72	1,14	0,2642	0,3729	0,1087	21,74
8	1,14	1,57	0,3729	0,4418	0,0689	13,78
9	1,57	∞	0,4418	0,5	0,0582	11,64
					$\sum p_i=1$	$\sum n_i'=200$

1. Por dos muestras independientes de volúmenes n_1 y n_2, extraídas de los conjuntos generales normales X e Y, se han encontrado las dispersiones muestrales corregidas s_X^2 y s_Y^2. Para el nivel de significación α verificar la hipótesis nula H_0: $D(X) = D(Y)$ sobre la igualdad de las dispersiones generales, siendo la hipótesis alternativa H_1: $D(X) > D(Y)$, si:
a) $n_1 = 21$, $n_2 = 16$, $s_X^2 = 3{,}6$, $s_Y^2 = 2{,}4$, $\alpha = 0{,}05$;
b) $n_1 = 13$, $n_2 = 18$, $s_X^2 = 0{,}72$, $s_Y^2 = 0{,}20$, $\alpha = 0{,}01$.

Respuesta a) $F_{obs} = 1{,}5$; $F_{cr}(0{,}05;\ 20;\ 15) = 2{,}33$.
No hay porque rechazar la hipótesis nula;
b) $F_{obs} = 3{,}6$; $F_{cr}(0{,}01;\ 12;\ 17) = 3{,}46$.
La hipótesis fundamental se rechaza.

2. Por dos muestras independientes de volúmenes n y m, escogidas de los conjuntos generales normales X e Y, se han encontrado las medias muestrales $\overline{x}$ e $\overline{y}$. Las dispersiones generales $D(X)$ y $D(Y)$ son conocidas. Para el nivel de significación α verificar la hipótesis fundamental H_0: $M(X) = M(Y)$ sobre la igualdad de las esperanzas matemáticas, para la hipótesis alternativa H_1: $M(X) \neq M(Y)$, si:
a) $n = 30$, $m = 20$, $D(X) = 120$, $D(Y) = 100$, $\alpha = 0{,}05$;
b) $n = 50$, $m = 40$, $D(X) = 50$, $D(Y) = 120$, $\alpha = 0{,}01$.

Respuesta a) $Z_{obs} = 1$, $z_{cr} = 1{,}96$.
No hay porque rechazar la hipótesis fundamental;
b) $Z_{obs} = 10$, $z_{cr} = 2{,}58$.
La hipótesis fundamental se rechaza.

3. Por dos muestras independientes de volúmenes $n = 5$ y $m = 6$, escogidas de los conjuntos generales normales X e Y, se han encontrado las medias muestrales $\overline{x} = 15{,}9$, $\overline{y} = 14{,}1$ y las dispersiones muestrales corregidas $s_X^2 = 14{,}76$, $s_Y^2 = 4{,}92$. Para el nivel de significación 0,05 verificar la hipótesis fundamental H_0: $M(X) = M(Y)$ de igualdad de las esperanzas matemáticas, siendo la hipótesis alternativa H_1: $M(X) \neq M(Y)$.

Advertencia. Comparar previamente las dispersiones.

Respuesta $T_{obs} = 0{,}88$, $t_{cr}(0{,}05;\ 9) = 2{,}26$.
No hay porque rechazar la hipótesis fundamental.

4. De un conjunto general normal con desviación cuadrática media conocida $\sigma = 2{,}1$ se ha escogido la muestra de volumen $n = 49$ y por ella se ha encontrado la media muestral $\overline{x} = 4{,}5$. Para el nivel de significación 0,05 hay que verificar la hipótesis fundamental H_0: $a = 3$ sobre la igualdad entre la esperanza matemática y el valor hipotético, si la hipótesis alternativa H_1: $a \neq 3$.

Respuesta $U_{obs} = 5$, $u_{cr} = 1{,}96$.
La hipótesis fundamental se rechaza.

5. Por la muestra de volumen $n = 16$, escogida de un conjunto general normal, se han encontrado la media muestral $\overline{x} = 12{,}4$ y la desviación cuadrática media «corregida» $s = 1{,}2$. Para el nivel de sig-

nificación 0,05 verificar la hipótesis nula H_0: $a = 11,8$ sobre la igualdad entre la esperanza matemática y el valor hipotético, siendo la hipótesis alternativa H_1: $a \neq 11,8$.

Respuesta. $T_{obs} = 2$, t_{cr} (0,05; 15) = 2,13.
No hay porque rechazar la hipótesis fudamental.

6. Con dos aparatos se han medido 5 piezas y se han obtenido los siguientes resultados (en mm):

$$x_1=4, \quad x_2=5, \quad x_3=6, \quad x_4=7, \quad x_5=8;$$

$$y_1=5, \quad y_2=5, \quad y_3=9, \quad y_4=4, \quad y_5=6;$$

Verificar cómo se diferencian, de manera significativa o insignificativa, los resultados de las mediciones, para el nivel de significación 0,05.

Respuesta $T_{obs} = 10,54$, t_{cr} (0,05; 4) = 2,78.
La diferencia de los resultados de las mediciones es significativa.

7. Por 100 experimentos independientes se ha encontrado la frecuencia relativa $\frac{m}{n} = 0,15$. Verificar para el nivel de significación la hipótesis fundamental H_0: $p = 0,17$ sobre la igualdad entre la frecuencia relativa y la probabilidad hipotética, si la hipótesis alternativa H_1: $p \neq 0,17$.

Respuesta $|U_{obs}| = 0,53$, $u_{cr} = 1,96$.
No hay porque rechazar la hipótesis fundamental.

8. Por cinco muestras independientes de volúmenes $n_1 = 7$, $n_2 = 9$, $n_3 = 10$, $n_4 = 12$, $n_5 = 12$, escogidas de conjuntos generales normales, se han encontrado las dispersiones muestrales: 0,27; 0,32; 0,40; 0,42; 0,48. Para el nivel de significación 0,05 conviene verificar la hipótesis nula sobre la homogeneidad de las dispersiones (la región crítica es de derecha).

Advertencia. Aplicar el criterio de Bartlett (§ 19).

Respuesta $V = 6,63$, χ^2_{cr} (0,05; 4) = 9,5. No hay porque rechazar la hipótesis fundamental.

9. Por cuatro muestras independientes de igual volumen $n = 17$ extraídas de conjuntos normales, se han hallado las dispersiones muestrales corregidas: 2,12; 2,32; 3,24; 4,32. Se necesita: a) para el nivel de significación 0,05, verificar la hipótesis fundamental de igualdad de las dispersiones generales (la región crítica es de derecha), b) estimar la dispersión general.

Advertencia. Utilizar el criterio de Cochran (§ 20).

Respuesta a) $G_{obs} = 0,36$, G_{cr} (0,05; 16,4) = 0,4366.
No hay porque rechazar la hipótesis fundamental; b) $\sigma = 3$.

10. Por un muestra de volumen $n = 62$, escogida de un conjunto normal bidimensional (X, Y), está hallado el coeficiente de correlación muestral $r_m = 0,6$. Para el nivel de significación 0,05 verificar la hipótesis nula H_0: $r_g = 0$ de igualdad a cero del coeficiente de correlación general, siendo la hipótesis alternativa $r_g \neq 0$.

Respuesta $T_{obs} = 5,81$, t_{cr} (0,05; 60) = 2,0.
La hipótesis nula se rechaza.

11. Para el nivel de significación 0,05, verificar la hipótesis de la distribución normal de un conjunto general si son conocidas las frecuencias empíricas y teóricas:

a) frecuencias empír.	6	12	16	40	13	8	5			
frecuencias teór.	4	11	15	43	15	6	6;			
b) frecuencias empír.	5	6	14	32	43	39	30	20	6	8;
frecuencias teór.	4	7	12	29	48	35	34	18	7	6
c) frecuencias empír.	5	13	12	44	8	12	6			
frecuencias teór.	2	20	12	35	15	10	6.			

Respuesta a) $\chi^2_{obs} = 2{,}5$, $\chi^2_{cr}(0{,}05;\ 4) = 9{,}5$.

No hay porque rechazar la hipótesis;

b) $\chi^2_{obs} = 3$, $\chi^2_{cr}(0{,}05;\ 7) = 14{,}1$.

No hay porque rechazar la hipótesis;

c) $\chi^2_{obs} = 13$, $\chi^2_{cr}(0{,}05;\ 4) = 9{,}5$.

La hipótesis se rechaza.

Capítulo veinte

ANALISIS DE DISPERSION DE UN FACTOR

§ 1. Comparación de varias medias. Noción de análisis de dispersión

Supongamos que los conjuntos generales $X_1, X_2, \ldots, X_p$ están distribuidos normalmente y tienen igual dispersión, aunque desconocida; las esperanzas matemáticas también son desconocidas, pero pueden ser diferentes. Para un nivel de significación dado hay que verificar por la media muestral la hipótesis nula (fundamental)

$$H_0 : M(X_1) = M(X_2) = \ldots = M(X_p)$$

sobre la igualdad de todas las esperanzas matemáticas. En otras palabras, hay que establecer cómo se diferencian, de manera significativa o insignificativa, las medias muestrales Podría pensarse que para comparar varias medias ($p > 2$) se pueden cotejar de dos en dos. Sin embargo, al crecer el número de medias, aumenta también la diferencia máxima entre ellas: la media de una nueva muestra puede resultar mayor que la máxima o la mínima de las medias obtenidas antes del nuevo experimento. Por esta causa, para comparar varias medias se utiliza otro método, basado en la comparación de las dispersiones y, por eso, llamada *análisis de dis-*

persión (desarrollada fundamentalmente por el estadístico inglés R. Fisher).

El análisis de dispersión se utiliza en la práctica para establecer si cierto factor *cualitativo* F que tiene p niveles $F_1, F_2, \ldots, F_p$, ejerce una influencia considerable en la magnitud que se estudia X. Por ejemplo, si se necesita aclarar qué tipo de abono es más eficaz para obtener una cosecha máxima, tendremos que el factor F es el abono, y su nivel, el tipo de abono.

La idea fundamental del análisis de dispersión es la comparación de la «dispersión de factor» ocasionada por la acción de un factor y la «dispersión residual» debida a causas fortuitas. Si la diferencia entre estas dispersiones es significativa, el factor ejerce una influencia considerable sobre X; en este caso, los valores medios observados en cada nivel (medias de grupo) se diferenciarán también de manera significativa.

Si ya se ha establecido que el factor influye esencialmente en X, y se necesita aclarar, cúal de los niveles ejerce una acción máxima, se hace una comparación suplementaria de dos en dos de las medias.

A veces el análisis de dispersión se utiliza para establecer la *homogeneidad* de varios conjuntos (las dispersiones de estos conjuntos son iguales por hipótesis; si el análisis de dispersión indica que las esperanzas matemáticas son idénticas, entonces los conjuntos son homogéneos). Los conjuntos homogéneos se pueden reunir en uno y con ello obtener respecto a éste una información más completa, y, por lo tanto, conclusiones más fiables.

En casos más complejos, se estudia la acción de varios factores sobre algunos niveles constantes o casuales y se aclara la influencia de niveles individuales y sus combinaciones (*análisis de múltiples factores*).

Nos limitaremos al caso simple del análisis de un factor, cuando sobre X actúa solamente un factor que tiene p niveles constantes.

§ 2. Sumas total, de factor y residual de los cuadrados de las desviaciones

Supongamos que sobre el carácter cuantitativo normalmente distribuido X actúa el factor F que tiene p niveles constantes. Vamos a admitir que el número de observaciones en cada nivel es idéntico e igual a q.

Supongamos que se han observado pq valores de x_{ij} del carácter X, donde i es el número de experimentos ($i = 1, 2, \ldots, q$), j es el número de niveles del factor ($j = 1, 2, \ldots, \ldots, p$). Los resultados de las observaciones están expuestas en la tabla 30.

Tabla 30

Número de experimento	Niveles del factor F_j			
	F_1	F_2	...	F_p
1	x_{11}	x_{12}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
...	...	...	...	...
q	x_{q1}	x_{q2}	...	x_{qp}
Media de grupo	$\bar{x}_{gr1}$	$\bar{x}_{g2r}$	...	$\bar{x}_{grp}$

Introducimos según la definición: $S_{\text{tot}} = \sum_{j=1}^{p} \sum_{i=1}^{q} (x_{ij} - \bar{x})^2$ (*suma total* de los cuadrados de las desviaciones de los valores observados respecto de la media general $\bar{x}$),

$$S_{\text{fac}} = q \sum_{j=1}^{p} (\bar{x}_{\text{gr}\,j} - \bar{x})^2$$

(*suma de factor* de los cuadrados de las desviaciones de las medias de grupos respecto de la media general que caracteriza la dispersión «entre grupos»),

$$S_{\text{res}} = \sum_{i=1}^{q} (x_{i1} - \bar{x}_{\text{gr}1})^2 + \sum_{i=1}^{q} (x_{i2} - \bar{x}_{\text{gr}2})^2 + \ldots +$$

$$+ \sum_{i=1}^{q} (x_{ip} - \bar{x}_{\text{gr}\,p})^2$$

(*suma residual* de los cuadrados de las desviaciones de los valores observados del grupo respecto a su media de grupo que caracteriza la dispersión «dentro de los grupos»).

La suma residual se halla prácticamente por la igualdad (§ 3, corolario):

$$S_{\text{res}} = S_{\text{tot}} - S_{\text{fac}}.$$

Mediante transformaciones elementales es posible obtener fórmulas más convenientes para los cálculos:

$$S_{tot} = \sum_{j=1}^{p} P_j - \frac{\left[\sum_{j=1}^{p} R_j\right]^2}{pq}, (*) \quad S_{fac} = \frac{\sum_{j=1}^{p} R_j^2}{q} - \frac{\left[\sum_{j=1}^{p} R_j\right]^2}{pq}, \quad (**)$$

donde $P_j = \sum_{i=1}^{q} x_{ij}^2$ es la suma de los cuadrados de los valores del carácter en el nivel F_j,

$R_j = \sum_{i=1}^{q} x_{ij}$ es la suma de los valores del carácter en el nivel F_j.

Nota. Para simplificar los cálculos se resta de cada valor observado un mismo número C aproximadamente igual a la media general. Si los valores reducidos $y_{ij} = x_{ij} - C$, tendremos que

$$S_{tot} = \sum_{j=1}^{p} Q_j - \frac{\left[\sum_{j=1}^{p} T_j\right]^2}{pq}, (***) \quad S_{fac} = \frac{\sum_{j=1}^{p} T_j^2}{q} - \frac{\left[\sum_{j=1}^{p} T_j\right]^2}{pq}, (****)$$

donde $Q_j = \sum_{i=1}^{q} y_{ij}^2$ es la suma de los cuadrados de los valores reducidos del carácter en el nivel F_j.

$T_j = \sum_{i=1}^{q} y_{ij}$ es la suma de los valores reducidos del carácter en el nivel F_j.

Para deducir las fórmulas (***) y (****) es suficiente poner $x_{ij} = y_{ij} + C$ en la correlación (*) y $R_j = \sum_{i=1}^{q} x_{ij} = \sum_{i=1}^{q} (y_{ij} + C) = \sum_{i=1}^{q} y_{ij} + qC = T_j + qC$ en la correlación (**).

Explicaciones.

1. Cerciorémonos de que S_{fac} caracteriza la acción del factor F. Admitamos que el factor ejerce una influencia considerable sobre X. En tal caso, el grupo de valores observados del carácter en un nivel determinado, será, en general, dife-

rente de los grupos de observaciones en otros niveles. Por lo tanto, también se diferenciarán las medias de grupos; además, cuanto más están dispersos alrededor de la media general, tanto mayor es la acción del factor. De aquí se deduce que para estimar la acción del factor conviene sumar los cuadrados de las desviaciones de las medias de grupos respecto de la media general (la desviación se eleva al cuadrado para excluir la eliminación de las desviaciones positivas y negativas). Multiplicando esta suma por q obtenemos S_{fac}. Así pues, S_{fac} caracteriza la acción del factor.

2. Demostremos que S_{res} refleja la influencia de causas fortuitas. Aparentemente las observaciones de un grupo no deben diferenciarse. Sin embargo, puesto que sobre X, además del factor F, actúan tambien causas fortuitas, las observaciones de un mismo grupo, en general, son distintas y, por consiguiente, están dispersas alrededor de su media de grupo. De aquí se deduce que para estimar la influencia de las causas fortuitas conviene sumar los cuadrados de las desviaciones de los valores observados de cada grupo respecto de su media de grupo, es decir, S_{res}. De este modo, S_{res} caracteriza la acción de las causas fortuitas.

3. Demostremos que S_{tot} refleja la influencia tanto del factor como de las causas fortuitas. Vamos a examinar todas las observaciones como un conjunto único. Los valores observados del carácter son distintos debido a la acción del factor y de las causas fortuitas. Para estimar esta acción conviene sumar los cuadrados de las desviaciones de los valores observados respecto de la media general, es decir, S_{tot}.

De este modo, S_{tot} caracteriza la influencia del factor y de las causas fortuitas.

Veamos un ejemplo que nos muestra claramente que la suma de factor refleja la influencia del factor, y la residual, la influencia de las causas fortuitas.

Ejemplo. Mediante dos aparatos se han realizado de a 2 mediciones de una magnitud física, cuya dimensión verdadera es igual a x. Considerando como factor el error sistemático C, y como sus niveles, los errores sistemáticos C_1 y C_2 respectivamente del primer y segundo aparato, conviene demostrar que S_{fac} se determina por los errores sistemáticos, mientras que S_{res}, por los errores aleatorios de las mediciones.

SOLUCION. Introducimos las designaciones:

α_1, α_2, errores aleatorios de la primera y segunda mediciones con el primer aparato,

β_1, β_2, errores aleatorios de la primera y segunda mediciones con el segundo aparato.

En tal caso, los valores observados de los resultados de las mediciones, respectivamente, son iguales a (el primer subíndice de x indica el número de medición, y el segundo, el número del aparato):

$$x_{11} = x + C_1 + \alpha_1, \quad x_{21} = x + C_1 + \alpha_2;$$
$$x_{12} = x + C_2 + \beta_1, \quad x_{22} = x + C_2 + \beta_2.$$

Los valores medios de las mediciones del primer y segundo aparatos respectivamente son iguales a:

$$\bar{x}_{gr1} = x + C_1 + \frac{\alpha_1 + \alpha_2}{2} = x + C_1 + \alpha,$$
$$\bar{x}_{gr2} = x + C_2 + \frac{\beta_1 + \beta_2}{2} = x + C_2 + \beta.$$

La media general es

$$\bar{x} = \frac{\bar{x}_{gr1} + \bar{x}_{gr2}}{2} = x + \frac{C_1 + C_2}{2} + \frac{\alpha + \beta}{2},$$

la suma de factor es

$$S_{fac} = (\bar{x}_{gr1} - \bar{x})^2 + (\bar{x}_{gr2} - \bar{x})^2.$$

Poniendo las magnitudes encerradas entre paréntesis, después de transformaciones elementales, obtenemos

$$S_{fac} = \frac{(C_1 - C_2)^2}{2} + (C_1 - C_2)(\alpha - \beta) + \frac{(\alpha - \beta)^2}{2}.$$

Como vemos, S_{fac} se determina, principalmente, por el primer sumando (puesto que los errores aleatorios de las mediciones son pequeños) y, por lo tanto, refleja realmente la influencia del factor C.

La suma residual es

$$S_{res} = (x_{11} - \bar{x}_{gr1})^2 + (x_{21} - \bar{x}_{gr1})^2 + (x_{12} - \bar{x}_{gr2})^2 + (x_{22} - \bar{x}_{gr2})^2.$$

Sustituyendo las magnitudes encerradas entre paréntesis obtenemos

$$S_{res} = [(\alpha_1 - \alpha)^2 + (\alpha_2 - \alpha)^2] + [(\beta_1 - \beta)^2 + (\beta_2 - \beta)^2].$$

Vemos que S_{res} se determina por los errores aleatorios de las mediciones y, por lo tanto, refleja realmente la influencia de las causas fortuitas.

Nota. El hecho de que S_{res} se debe a causas fortuitas también se deduce de la igualdad (§ 3, corolario)

$$S_{res} = S_{tot} - S_{fac}.$$

En efecto, S_{tot} es el resultado de la acción del factor y de las causas fortuitas; restando S_{fac} excluimos la influencia del factor. Por lo tanto, «la parte restante» refleja la influencia de las causas fortuitas.

§ 3. Vínculo entre las sumas total, de factor y residual

Demostraremos que

$$S_{tot} = S_{fac} + S_{res}.$$

Para simplificar la deducción nos limitaremos a dos niveles ($p = 2$) y dos experimentos en cada nivel ($q = 2$). Los resultados de las pruebas las presentamos en la forma de la tabla 31.

Tabla 31

Número de la prueba i	Niveles del factor F_j	
	F_1	F_2
1	x_{11}	x_{12}
2	x_{21}	x_{22}
$\overline{x}_{grj}$	$\overline{x}_{gr1}$	$\overline{x}_{gr2}$

En tal caso,

$$S_{tot} = (x_{11} - \overline{x})^2 + (x_{21} - \overline{x})^2 + (x_{12} - \overline{x})^2 + (x_{22} - \overline{x})^2.$$

Restamos y sumamos a cada valor observado en el primer nivel la media de grupo $\overline{x}_{gr1}$, y en el segundo, $\overline{x}_{gr2}$.

Elevando al cuadrado y teniendo en cuenta que la suma de todos los productos duplicados es igual a cero (recomendamos al lector verificar esto individualmente), obtenemos

$$S_{tot} = 2[(\overline{x}_{gr1} - \overline{x})^2 + (\overline{x}_{gr2} - x)^2] + [(x_{11} - \overline{x}_{gr1})^2 + \\ + (x_{21} - \overline{x}_{gr1})^2 + (x_{12} - \overline{x}_{gr2})^2 + (x_{22} - \overline{x}_{gr2})^2] = S_{fac} + S_{res}.$$

Así pues, $S_{tot} = S_{fac} + S_{rec}.$

Corolario. De la igualdad obtenida se deduce un importante corolario:

$$S_{res} = S_{tot} - S_{fac}.$$

De aquí se aprecia que no hay necesidad de calcular directamente la suma residual; es suficiente hallar las sumas total y de factor, y luego su diferencia.

§ 4. Dispersiones total, de factor y residual

Dividiendo las sumas de los cuadrados de las desviaciones por el correspondiente número de grados de libertad, obtenemos las dispersiones total, de factor y residual:

$$s^2_{tot} = \frac{S_{tot}}{pq-1}, \quad s^2_{fac} = \frac{S_{fac}}{p-1}, \quad s^2_{res} = \frac{S_{res}}{p(q-1)},$$

donde p es el número de niveles del factor,

q es el número de observaciones en cada nivel.

Si la hipótesis fundamental sobre la igualdad de las medias es cierta, todas estas dispersiones son estimaciones no desviadas de la dispersión general. Por ejemplo, teniendo en cuenta que el volumen de la muestra es $n = pq$, deducimos que $s^2_{tot} = \frac{S_{tot}}{pq-1} = \frac{S_{tot}}{n-1}$, es la dispersión muestral corregida, que se sabe, es la estimación no desviada de la dispersión general.

Nota. El número de grados de libertad $p(q-1)$ de la dispersión residual es igual a la diferencia entre los números de grados de libertad de las dispersiones total y de factor. En efecto,

$$(pq-1) - (p-1) = pq - p = p(q-1).$$

§ 5. Comparación de varias medias por el método de análisis de dispersión

Volvamos al problema planteado en el § 1: verificar para un nivel de significación prefijado la hipótesis nula sobre la igualdad de varias medias ($p > 2$) de conjuntos normales con dispersiones desconocidas, pero idénticas. Demostramos que la resolución de este problema se reduce a comparar las dispersiones de factor y residual según el criterio de Fisher—Snedecor (cap. XIX, § 8).

1. Supongamos que la hipótesis nula sobre la igualdad de varias medias (en adelante las llamaremos de grupos) es

cierta. En este caso, las dispersiones de factor y residual son estimaciones no desviadas de la dispersión general desconocida (§ 4) y, por lo tanto, se diferencian de manera significativa. Si se comparan estas estimaciones por el criterio F, evidentemete el criterio indica que la hipótesis nula sobre la igualdad entre las dispersiones de factor y residual debe aceptarse.

Por consiguiente, si la hipótesis sobre la igualdad de las medias de grupos es cierta, es cierta también la hipótesis de la igualdad de las dispersiones de factor y residual.

2. Supongamos que la hipótesis nula sobre la igualdad entre las medias de grupos es falsa. En este caso, al aumentar la divergencia entre las medias de grupos se incrementará la dispersión de factor, y con ella también la relación $F_{obs} = \frac{S^2_{fac}}{S^2_{res}}$. En suma F_{obs} resulta mayor que F_{cr}, y por lo tanto, la hipótesis de la igualdad de dispersiones será rechazada.

De este modo, si la hipótesis de la igualdad de las medias de grupos es falsa, también es falsa la hipótesis de la igualdad entre las dispersiones de factor y residual.

Se demuestra fácilmente a la inversa la certeza de las tesis inversas: de la justeza (falsedad) de la hipótesis de las dispersiones se deduce la justeza (falsedad) de la hipótesis de las medias.

Así, *para verificar la hipótesis nula de la igualdad de las medias de grupos de conjuntos normales con idénticas dispersiones, es suficiente verificar por el criterio F la hipótesis nula sobre la igualdad de las dispersiones de factor y residual.* En esto consiste, precisamente, el método de análisis de dispersión.

Nota 1. Si la dispersión de factor resulta menor que la residual, ya de aquí se desprende la certeza de la hipótesis de la igualdad de las medias de grupos y, por eso, no hay necesidad de recurrir al criterio F.

Nota. 2. Si no se está seguro de la certeza de la hipótesis de la igualdad de las dispersiones de los conjuntos examinados p, esta hipótesis debej verificarse, por ejemplo, por el criterio de Cochran.

Ejemplo. Se han realizado 4 experimentos en cada uno de los tres niveles. Los resultados de estos experimentos están expuestos en la tabla 32. Para el nivel de significación 0,05 hay que verificar por el método de análisis de dispersión la hipótesis nula sobre la igualdad de las medias de grupos. Se supone que las muestras se han escogido de los conjuntos normales con idénticas dispersiones.

Número del experimento i	Niveles del factor F_j		
	F_1	F_2	F_3
1	51	52	42
2	52	54	44
3	56	56	50
4	57	58	52
$\bar{x}_{grj}$	54	55	47

SOLUCION. Para simplificar el cálculo restamos $C = 52$ de cada valor observado: $y_{ij} = x_{ij} - 52$. Formamos la tabla de cálculo 33.

Tabla 33

Número del experimento i	Niveles del factor F_j						
	F_1		F_2		F_3		
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	
1	−1	1	0	0	−10	100	
2	0	0	2	4	−8	64	
3	4	16	4	16	−2	4	
4	5	25	6	36	0	0	
$S_j = \sum_{i=1}^{4} y_{ij}^2$		42		56		168	$\sum S_j = 266$
T_j	8		12		−20		$\sum T_j = 0$
T_j^2	64		144		400		$\sum T_j^2 = 608$

Utilizando la tabla y teniendo en cuenta que el número de niveles del factor $p = 3$, el número de experimentos en cada nivel $q = 4$, hallamos las sumas total y de factor de los cuadrados de las desviaciones [§ 2, fórmulas (***) y (****)]:

$$S_{tot} = \sum_{j=1}^{p} S_j - \frac{[\sum_{j=1}^{p} T_j]^2}{pq} = 266 - 0 = 266;$$

$$S_{fac} = \frac{\sum_{j=1}^{p} T_j^2}{q} - \frac{[\sum_{j=1}^{p} T_j]^2}{pq} = \frac{608}{4} - 0 = 152.$$

Hallamos la suma residual de los cuadrados de las desviaciones:

$$S_{res} = S_{tot} - S_{fac} = 266 - 152 = 114.$$

Hallamos las dispersiones de factor y residual:

$$s_{fac}^2 = \frac{S_{fac}}{p-1} = \frac{152}{3-1} = 76;$$

$$s_{res}^2 = \frac{S_{res}}{p(q-1)} = \frac{114}{3(4-1)} = \frac{114}{9} = 12{,}67.$$

Comparamos las dispersiones de factor y residual por el criterio F (cap. XIX, § 8), para lo cual hallamos el valor observado del criterio:

$$F_{obs} = \frac{s_{fac}^2}{s_{res}^2} = \frac{76}{12{,}67} = 6.$$

Tomando en consideración que el número de grados de libertad del numerador $k_1 = 2$, y del denominador $k_2 = 9$, así como el nivel de significación $\alpha = 0{,}05$, por la tabla hallamos el punto crítico $F_{cr}(0{,}05; 2; 9) = 4{,}26$.

Puesto que $F_{obs} > F_{cr}$, la hipótesis nula sobre la igualdad de las medias de grupos se rechaza. En otras palabras, las medias de grupos «en total» se diferencian de modo significativo. Si es necesario comparar las medias de dos en dos, debe utilizarse el criterio t de Student.

Nota 3. Si los valores observados x_{ij} son fracciones decimales con una cifra después de la coma, conviene pasar a los números $y_{ij} = 10 x_{ij} - C$, donde C es aproximadamente el valor medio del número $10 x_{ij}$. En resumen obtenemos números *enteros* relativiamente pequeños. Aunque en este caso las dispersiones de factor y residual aumentan 10^2 veces, su relación no varia. Por ejemplo, si $x_{11} = 12{,}1$, $x_{21} = 12{,}2$, $x_{31} = 12{,}6$, tomando $y_{ij} = 10 \cdot x_{ij} - 123$, obtenemos: $y_{11} = 121 - 123 = -2$, $y_{21} = 122 - 123 = -1$, $y_{31} = 126 - 123 = 3$.

Se procede análogamente, si después de la coma se tiene k cifras:

$$y_{ij} = 10^k \cdot x_{ij} - C.$$

Problemas

En los problemas 1—2 se necesita verificar para el nivel de significación 0,05 la hipótesis nula sobre la igualdad de las medias de grupos. Se supone que las muestras han sido extraídas de conjuntos normales con idénticas dispersiones generales.

1.

Número del experimento i	Niveles del factor F_j				
	F_1	F_2	F_3	F_4	F_5
1	42	66	35	64	70
2	55	91	50	70	79
3	67	96	60	79	88
4	67	98	69	81	90
$\overline{x}_{grj}$	57,75	87,75	53,50	73,50	81,75

Respuesta $F_{obs}=6,13$; $F_{cr}(0,05;\ 4;\ 15)=3,06$.
La hipótesis nula se rechaza.

2.

Número del experimento i	Niveles del factor F_j			
	F_1	F_2	F_3	F_4
1	6	6	9	7
2	7	7	12	9
3	8	11	13	10
4	11	12	14	10
$\overline{x}_{grj}$	8	9	12	9

Respuesta $F_{obs}=2,4$; $F_{cr}(0,05;\ 3;\ 12)=3,49$.
No hay porque rechazar la hipótesis nula.

Tabla de valores de la función $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025

x	0	1	2	3	4	5	6	7	8	9
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Suplemento 2

Tabla de valores de la función $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz.$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,0000	0,18	0,0714	0,36	0,1406	0,54	0,2054
0,01	0,0040	0,19	0,0753	0,37	0,1443	0,55	0,2088
0,02	0,0080	0,20	0,0793	0,38	0,1480	0,56	0,2123
0,03	0,0120	0,21	0,0832	0,39	0,1517	0,57	0,2157
0,04	0,0160	0,22	0,0871	0,40	0,1554	0,58	0,2190
0,05	0,0199	0,23	0,0910	0,41	0,1591	0,59	0,2224
0,06	0,0239	0,24	0,0948	0,42	0,1628	0,60	0,2257
0,07	0,0279	0,25	0,0987	0,43	0,1664	0,61	0,2291
0,08	0,0319	0,26	0,1026	0,44	0,1700	0,62	0,2324
0,09	0,0359	0,27	0,1064	0,45	0,1736	0,63	0,2357
0,10	0,0398	0,28	0,1103	0,46	0,1772	0,64	0,2389
0,11	0,0438	0,29	0,1141	0,47	0,1808	0,65	0,2422
0,12	0,0478	0,30	0,1179	0,48	0,1844	0,66	0,2454
0,13	0,0517	0,31	0,1217	0,49	0,1879	0,67	0,2486
0,14	0,0557	0,32	0,1255	0,50	0,1915	0,68	0,2517
0,15	0,0596	0,33	0,1293	0,51	0,1950	0,69	0,2549
0,16	0,0636	0,34	0,1331	0,52	0,1985	0,70	0,2580
0,17	0,0675	0,35	0,1368	0,53	0,2019	0,71	0,2611

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,72	0,2642	1,09	0,3621	1,46	0,4279	1,83	0,4664
0,73	0,2673	1,10	0,3643	1,47	0,4292	1,84	0,4671
0,74	0,2703	1,11	0,3665	1,48	0,4306	1,85	0,4678
0,75	0,2734	1,12	0,3686	1,49	0,4319	1,86	0,4686
0,76	0,2764	1,13	0,3708	1,50	0,4332	1,87	0,4693
0,77	0,2794	1,14	0,3729	1,51	0,4345	1,88	0,4699
0,78	0,2823	1,15	0,3749	1,52	0,4357	1,89	0,4706
0,79	0,2852	1,16	0,3770	1,53	0,4370	1,90	0,4713
0,80	0,2881	1,17	0,3790	1,54	0,4382	1,91	0,4719
0,81	0,2910	1,18	0,3810	1,55	0,4394	1,92	0,4726
0,82	0,2939	1,19	0,3830	1,56	0,4406	1,93	0,4732
0,83	0,2967	1,20	0,3849	1,57	0,4418	1,94	0,4738
0,84	0,2995	1,21	0,3869	1,58	0,4429	1,95	0,4744
0,85	0,3023	1,22	0,3883	1,59	0,4441	1,96	0,4750
0,86	0,3051	1,23	0,3907	1,60	0,4452	1,97	0,4756
0,87	0,3078	1,24	0,3925	1,61	0,4463	1,98	0,4761
0,88	0,3106	1,25	0,3944	1,62	0,4474	1,99	0,4767
0,89	0,3133	1,26	0,3962	1,63	0,4484	2,00	0,4772
0,90	0,3159	1,27	0,3980	1,64	0,4495	2,02	0,4783
0,91	0,3186	1,28	0,3997	1,65	0,4505	2,04	0,4793
0,92	0,3212	1,29	0,4015	1,66	0,4515	2,06	0,4803
0,93	0,3238	1,30	0,4032	1,67	0,4525	2,08	0,4812
0,94	0,3264	1,31	0,4049	1,68	0,4535	2,10	0,4821
0,95	0,3289	1,32	0,4066	1,69	0,4545	2,12	0,4830
0,96	0,3315	1,33	0,4082	1,70	0,4554	2,14	0,4838
0,97	0,3340	1,34	0,4099	1,71	0,4564	2,16	0,4846
0,98	0,3365	1,35	0,4115	1,72	0,4573	2,18	0,4854
0,99	0,3389	1,36	0,4131	1,73	0,4582	2,20	0,4861
1,00	0,3413	1,37	0,4147	1,74	0,4591	2,22	0,4868
1,01	0,3438	1,38	0,4162	1,75	0,4599	2,24	0,4875
1,02	0,3461	1,39	0,4177	1,76	0,4608	2,26	0,4881
1,03	0,3485	1,40	0,4192	1,77	0,4616	2,28	0,4887
1,04	0,3508	1,41	0,4207	1,78	0,4625	2,30	0,4893
1,05	0,3531	1,42	0,4222	1,79	0,4633	2,32	0,4898
1,06	0,3554	1,43	0,4236	1,80	0,4641	2,34	0,4904
1,07	0,3577	1,44	0,4251	1,81	0,4649	2,36	0,4909
1,08	0,3599	1,45	0,4265	1,82	0,4656	2,38	0,4913

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2,40	0,4918	2,60	0,4953	2,80	0,4974	2,98	0,4986
2,42	0,4922	2,62	0,4956	2,82	0,4976	3,00	0,49865
2,44	0,4927	2,64	0,4959	2,84	0,4977	3,20	0,49931
2,46	0,4931	2,66	0,4961	2,86	0,4979	3,40	0,49966
2,48	0,4934	2,68	0,4963	2,88	0,4980	3,60	0,499841
2,50	0,4938	2,70	0,4965	2,90	0,4981	3,80	0,499928
2,52	0,4941	2,72	0,4967	2,92	0,4982	4,00	0,499968
2,54	0,4945	2,74	0,4969	2,94	0,4984	4,50	0,499997
2,56	0,4948	2,76	0,4971	2,96	0,4985	5,00	0,499997
2,58	0,4951	2,78	0,4973				

Suplemento 3

Tabla de valores $t_{\gamma} = t(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,729	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Tabla de valores $q = q(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,221
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

Puntos críticos de la distribución χ^2

Número de grados de libertad k	Nivel de significación α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56

Número de grados de libertad k	Nivel de significación α					
	0,01	0,025	0,05	0,95	0,975	0,99
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Suplemento 6

Puntos críticos de la distribución t de Student

Número de grados de libertad k	Nivel de significación α (región crítica bilateral)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86

Número de grados de libertad k	Nivel de significación α (región crítica bilateral)					
	0,10	0,05	0,02	0,01	0,002	0,001
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Nivel de significación α (región crítica unilateral)					

Puntos críticos de la distribución F de Fisher — Snedecor

(k_1 — número de grados de libertad de la dispersión mayor)
(k_2 — número de grados de libertad de la dispersión menor)

Nivel de significación $\alpha = 0,01$

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	99,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Nivel de significación $\alpha = 0,05$

h_2 \ h_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,64	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Puntos críticos de la

(*k*—número de grados de liber

	Nivel de significa						
l \ *k*	1	2	3	4	5	6	7
2	0,9999	0,9950	0,9794	0,9586	0,9373	0,2172	0,8988
3	9933	9423	8831	8335	7933	7606	7335
4	9676	8643	7814	7212	6761	6410	6129
5	0,9269	0,7885	0,6957	0,6329	0,5875	0,5531	0,5259
6	8828	7218	6258	5635	5195	4866	4608
7	8376	6644	5685	5080	4659	4347	4105
8	0,7945	0,6152	0,5209	0,4627	0,4226	0,3932	0,3704
9	7544	5727	4810	4251	3870	3592	3378
10	7175	5358	4469	3934	3572	3308	3106
12	0,6528	0,4751	0,3919	0,3428	0,3099	0,2861	0,2680
15	5747	4069	3317	2882	2593	2386	2228
20	4799	3297	2654	2288	2048	1877	1748
24	0,4247	0,2871	0,2295	0,1970	0,1759	0,1608	0,1495
30	3632	2412	1913	1635	1454	1327	1232
40	2940	1915	1508	1281	1135	1033	0951
60	0,2151	0,1371	0,1069	0,0902	0,0796	0,0722	0,0668
120	1225	0759	0585	0489	0429	0387	0357
∞	0000	0000	0000	0000	0000	0000	0000

distribución de Cochran

tad; l — cantidad de muestras)

ción $\alpha = 0{,}01$						
8	9	10	16	36	144	∞
0,8823	0,8674	0,8539	0,7949	0,7067	0,6062	0,5000
7107	6912	6743	6059	5153	4230	3333
5897	5702	5536	4884	4057	3251	2500
0,5037	0,4854	0,4697	0,4094	0,3351	0,2644	0,2000
4401	4229	4084	3529	2858	2229	1667
3911	3751	3616	3105	2494	1929	1429
0,3522	0,3373	0,3248	0,2779	0,2214	0,1700	0,1250
3207	3067	2950	2514	1992	1521	1111
2945	2813	2704	2297	1811	1376	1000
0,2535	0,2419	0,2320	0,1961	0,1535	0,1157	0,0833
2104	2002	1918	1612	1251	0934	0667
1646	1567	1501	1248	0960	0709	0500
0,1406	0,1338	0,1283	0,1060	0,0810	0,0595	0,0417
1157	1100	1054	0867	0658	0480	0333
0898	0853	0816	0668	0503	0363	0250
0,0625	0,0594	0,0567	0,0461	0,0344	0,0245	0,0167
0334	0316	0302	0242	0178	0125	0083
0000	0000	0000	0000	0000	0000	0000

							Nivel de significa
l \ k	1	2	3	4	5	6	7
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332
3	9669	8709	7977	7457	7071	6771	6530
4	9065	7679	6841	6287	5895	5598	5365
5	0,8412	0,6838	0,5981	0,5440	0,5063	0,4783	0,4564
6	7808	6161	5321	4803	4447	4184	3980
7	7271	5612	4800	4307	3974	3726	3535
8	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185
9	6385	4775	4027	3584	3286	3067	2901
10	6020	4450	3733	3311	3029	2823	2666
12	0,5410	0,3924	0,3624	0,2880	0,2624	0,2439	0,2299
15	4709	3346	2758	2419	2195	2034	1911
20	3894	2705	2205	1921	1735	1602	1501
24	0,3434	0,2354	0,1907	0,1656	0,1493	0,1374	0,1286
30	2929	1980	1593	1377	1237	1137	1061
40	2370	1576	1259	1082	0968	0887	0827
60	0,1737	0,1131	0,0895	0,0765	0,0682	0,0623	0,0583
120	0998	0632	0495	0419	0371	0337	0312
∞	0000	0000	0000	0000	0000	0000	0000

ción α = 0,05

8	9	10	16	36	144	∞
0,8159	0,8010	0,7880	0,7341	0,6602	0,5813	0,5000
6333	6167	6025	5466	4748	4031	3333
5175	5017	4884	4366	3720	3093	2500
0,4387	0,4241	0,4118	0,3645	0,3066	0,2513	0,2000
3817	3682	3568	3135	2612	2119	1667
3384	3259	3154	2756	2278	1833	1429
0,3043	0,2926	0,2829	0,2462	0,2022	0,1616	0,1250
2768	2659	2568	2226	1820	1446	1111
2541	2439	2353	2032	1655	1308	1000
0,2187	0,2098	0,2020	0,1737	0,1403	0,1100	0,0833
1815	1736	1671	1429	1144	0889	0667
1422	1357	1303	1108	0879	0675	0500
0,1216	0,1160	0,1113	0,0942	0,0743	0,0567	0,0417
1002	0958	0921	0771	0604	0457	0333
0780	0745	0713	0595	0462	0347	0250
0,0552	0,0520	0,0497	0,0411	0,0316	0,0234	0,0167
0292	0279	0266	0218	0165	0120	0083
0000	0000	0000	0000	0000	0000	0000

INDICE ALFABÉTICO

A NUESTROS LECTORES:

«Mir» edita libros soviéticos traducidos al español, inglés, francés y árabe. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica; manuales para los centros de enseñanza superior y escuelas tecnológicas; literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia ficción.

Dirijan sus opiniones a Editorial MIR, I Rizhski per. 2, 129820 Moscú GSP I-110, URSS.

EN 1975 LA EDITORIAL MIR PUBLICARÁ:

CÁLCULO DE VARIACIONES

de M. Krasnov, G. Makarenko, A. Kiseliov

Los autores de este libro son Mijaíl Krasnov, Grigori Makarenko, candidatos a doctores en ciencias físico-matemáticas y docentes del Instituto Energético de Moscú, y Alexandr Kiseliov, colaborador científico superior del Instituto Unificado de Investigaciones Nucleares de la ciudad de Dubna.

Este compendio contiene problemas y ejercicios dedicados a ilustrar los diferentes principios de la teoría y los métodos de resolución de las ecuaciones por el cálculo de variaciones.

Al principio de cada capítulo se resumen los resultados principales, se exponen los conocimientos teóricos necesarios, las fórmulas requeridas y se estudian con gran detalle ejemplos típicos ilustrativos.

Este manual contiene más de 100 ejemplos analizados y 230 problemas destinados para resolverse independientemente. Unos problemas se acompañan con las respuestas, otros, con las referencias de cómo deben resolverse.

La obra está dirigida a los estudiantes de los centros de enseñanza técnica superior que se especializan en los cálculos matemáticos.

Formato 14,5 × 22 cm. Encuadernado en tela con sobrecubierta. 200 págs.

FUNDAMENTOS MATEMÁTICOS DE LA CIBERNÉTICA
de Yu. Kórshunov

El autor de este libro, doctor en ciencias técnicas, profesor Yuri Kórshunov, es Jefe de la Cátedra de Automática y Telemecánica del Instituto Radiotécnico.

En el manual se exponen a un nivel accesible para el ingeniero los métodos matemáticos básicos que han adquirido amplia difusión con motivo de la aparición y el empleo práctico de las calculadoras numéricas para fines de mando. Se han explicado y llevado hasta sus definiciones matemáticas formales muchísimos conceptos utilizados en la Cibernética. Se esclarecen los principios del enfoque cibernético para la resolución de las cuestiones de la optimización de los procesos de mando y se presentan los métodos matemáticos primordiales para solucionar dicho problema.

La exposición se acompaña de ejemplos y problemas que ayudan a entender los fundamentos teóricos principales.

Este libro está destinado para los estudiantes de facultades de automática y telemecánica y de las especialidades afines de los centros de enseñanza superior.